

Bush's science flashpoints

The president-elect seems to have failed to inspire scientists during his campaign. Although support for research will probably grow, some policies and appointments are likely to signal trouble ahead.

Incorrigible optimism is the primary characteristic of American society that differentiates it from all other countries. After last week's Supreme Court ruling awarded the presidency of the United States to George W. Bush, in less than auspicious circumstances, most Americans, of whatever political hue, are strongly inclined to give their president-elect the benefit of the doubt.

Some US citizens harbour more doubt than others, however. The gun lobby can have little doubt about its preferred candidate's aversion to gun control. In contrast, the minorities who believe themselves to have been disenfranchised in certain precincts of Florida may never see Bush as a legitimate president.

The scientific community, as far as one can generalize, sits between these two extremes. Many scientists have concerns about Bush's self-avowed lack of intellectual curiosity, and most probably voted against him — polls show him losing badly among Americans holding higher degrees. But these initial doubts need not undermine the relationship that must be built between science and the new administration. There are, however, some areas in which the view of the scientific community, as expressed through the relevant scientific societies, clearly falls into conflict with the likely agenda of the new administration (see pages 887 and 888).

The first of these potential flashpoints concerns the status of human embryonic stem-cell research at the National Institutes of Health (NIH). As *Nature* went to press, it was being suggested that the secretaryship of the Department of Health and Human Services, of which the NIH is part, might be awarded to someone who can be relied on, in effect, to oppose abortion on every available front.

It is to be fervently hoped that the directorship of the NIH itself will not be awarded on a similar basis. Although Harold Varmus, who led the agency for most of the Clinton years, was regarded by the Democrats as politically reliable, it was his exceptional scientific credentials that enabled the NIH to win impressive, bipartisan political support. The expansion of the NIH to a \$20 billion agency and beyond has become one of the most important and impressive projects of the US government. It requires a leader who enjoys not only the confidence of the incoming Republican administration, but also, by virtue of scientific stature and vision, that of the biomedical research community. A handful of such candidates exist.

A second potential flashpoint concerns the status of environmental science within the new administration. The application of this science to environmental regulation has always been highly charged

politically, and the Clinton administration certainly didn't make it any less so. The challenge for the new administration here is to rise above the temptation merely to avenge its political enemies on environmental issues and instead to prove to a sometimes sceptical public that Republicans value environmental protection. After all, that is what Richard Nixon was trying to do when he set up the Environmental Protection Agency.

It is also what Senator John McCain (Republican, Arizona) has been doing during his recent reconsideration of the global-warming issue, which was undertaken after the conservative senator ran his own race for the Republican nomination last winter and discovered on the campaign trail — amazingly enough — that young people are fearful for the future of the planet. Bush needs to take a leaf out of McCain's book and look again at the evidence for man-made global warming, and at America's possible role in averting it.

A third potential flashpoint concerns the Bush administration's pledge to deploy National Missile Defense. Many US physicists have hotly contested the technical claims made by the proponents of missile-defence systems. In the past, this particular issue has proven to be a disaster for relations between Republican administrations and scientists. But the stakes are not as high as they were during the Cold War. The forthcoming arguments about such matters as the effectiveness of decoys and the feasibility of modifications to the Anti-Ballistic Missile Treaty should be resolved without utterly souring relations between the White House and America's physicists.

In each of these areas, the role of a new science adviser to the president will be crucial. Bush may become the first president to place a biologist in this position, reflecting the gradual displacement of national security by health as the US government's highest scientific priority. The science and technology enterprise is one of the few areas of the government's business that enjoys genuine bipartisan support, and the new administration should build on the funding progress made by the outgoing administration and Congress, acting in unlikely concert.

The appointments Bush makes in the next few months will set the tone for his administration. Despite the alleged requirement for the new president to place Democrats in a few high-profile positions, it is the people further down, those who do the government's work, who will set the real tone. It will soon be apparent, from the scientific appointments alone, whether George W. seeks to unite his nation, or to divide it yet more deeply. ■

Futures' end

Given its stature and its heavy rotation on cable TV, it is a constant source of amazement that only 12 episodes of John Cleese's comedy *Fawlty Towers* were ever made. Wisely, the producers decided to quit while they were ahead, and every episode stands as a gem. Thus are legends made: live fast and die young.

In the same way, we and at least some readers have had fun with *Futures*, our science-fictional *jeu d'esprit*, which comes to a thumping

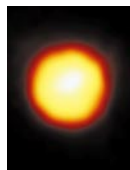
conclusion this week (see page 913). The series has covered the extinction and transfiguration of humanity; manifestations of divinity; conversations with androids, extraterrestrials and even bacteria; and the total destruction of the Earth (twice). As the inevitable end of the series approached, we began to ask ourselves what we'll be doing for excitement next. Like the Star Child at the conclusion of Arthur C. Clarke's millennial novel, *2001*, we're sure we'll think of something. ■







Divided opinion
Future of stem-cell research will be reviewed under Bush
p887



Probing question
Pluto set to reappear on NASA's list of priorities
p889



Light work
Fibre-optics research receives funding boost
p892



2000 in context
The science that made the headlines during the year
p894

Tussle starts for research posts as Bush rides into Washington

Tony Reichhardt and Paul Smaglik, Washington

Scientific leaders are anticipating continued strong support for research funding but potential fireworks on a range of science-related issues under the administration of President-elect George W. Bush.

"When there's a close balance of power, Republicans and Democrats tend to push those things they can agree upon, and that looks promising for science," says Peter Saundry, executive director of the National Council for Science and the Environment, which lobbies for environmental research. He adds that science "tends to do better under Republicans" because Democrats have other spending priorities.

Bush arrives in Washington this week to plan the administration that will take power at the end of January. Other than funding, the big question for science is to what degree the balance of power will shift towards the Republican party's right wing. If conservatives succeed in setting the agenda for the administration, researchers will soon find themselves embroiled in controversies ranging from stem-cell research to missile defence and global warming (see below and overleaf).

The speculative — but entertaining — game of guessing who will fill key positions in the Bush administration was well under way this week, although most of the top scientific positions in the administration will probably remain unfilled for several months.

Rita Colwell, director of the National Science Foundation, is the only senior science administrator likely to stay in place under Bush, as her position carries a six-year tenure that began in 1999.

Before serious attention turns to the directorship of the National Institutes of Health (NIH), Bush will have to nominate the NIH director's boss, the Secretary of Health and Human Services. Some lobbyists say that Gail Wilensky, who advised former President George Bush on health issues and now works at the humanitarian organization Project HOPE, is a front runner for the secretary's job. William Roper, former director of the Centers for Disease Control and Prevention, now dean of public health at the University of North Carolina at Chapel Hill, is also in the frame.

External candidates whose names are being floated for the NIH directorship include John Mendelsohn, president of the University of Texas M. D. Anderson Cancer Center, and Mike Brown, 1985 Nobel laureate in medicine who heads a genetics centre at the University of Texas Southwestern Medical Center. Tony Fauci, director of the National Institute of Allergy and Infectious Diseases, is seen as a viable internal candidate for the post.

NASA administrator Dan Goldin, having served an epic tenure since his 1992 appointment by Bush senior, is said to want to stay put. If Bush decides otherwise, names being put forward to replace him include Mark

Albrecht, an aerospace executive who led the National Space Council under Bush senior, and Michael Griffin, chief technical officer at the Orbital Sciences Corporation.

Marye Anne Fox, a chemist and chancellor of North Carolina State University, who advised Bush on science policy in Texas, is seen in Washington as a possible candidate to head the White House Office of Science and Technology Policy, which coordinates science policy across the government. The incumbent, physicist Neal Lane, announced last week that he is moving swiftly in the opposite direction to George W., and returning to a chair at Rice University in Texas. ■



At last: George W. Bush addresses the nation after his election as the next US president.

Stem-cell work in the balance

Matthew Davis, Washington

Conventional wisdom has it that George W. Bush's presidency could doom US researchers' efforts to obtain federal funding for work using human embryonic stem cells. But supporters of the research are optimistic that the new administration will at least review the issue before acting.

"We're very hopeful that the Bush administration will not restrict stem-cell research, and we will work toward that end," says Tim Leshan, director of public policy at the American Society for Cell Biology.

During the presidential campaign,

Bush's aides indicated that he would halt the National Institutes of Health (NIH) initiative to support the research. Although it was launched almost two years ago, this initiative has yet to fund any stem-cell research.

But supporters of the research think that neither Bush nor his closest advisers have yet given the debate much thought. They note that statements from Bush aides indicating his opposition to the NIH effort came in response to questions from a reporter, and were not official policy. And they are advising researchers to avoid a



Loss of support: stem-cell advocate Arlen Specter will lose reign over NIH budget.

► confrontational stance with the new administration, so as not to drive Bush into appeasing the opponents of embryo research.

The current NIH regime is still setting up a stem-cell research programme. This effort is rooted in a legal opinion, crafted under President Bill Clinton's administration, that says the US government can fund research with stem cells derived from human embryos, despite the agency being prohibited by law from supporting work with the embryos themselves. This opinion — which even some researchers think is shaky — would take little executive action to reverse.

Advocates of stem-cell research are worried that Bush could overturn the ruling without having to cut off funding for any research already underway. Rather, the administration could avoid the political heat that would come from ending the programme outright by requesting additional reviews that would, in effect, delay funding indefinitely.

Even if the current policy remained intact, NIH officials do not anticipate approving any embryonic stem-cell research until May at the earliest. The deadline for the first round of applications is 15 March.

People familiar with the process say that a key hold-up is the lack of embryonic stem-cell lines that comply with NIH guidelines. The NIH is also still vetting members of its new Human Pluripotent Stem Cell Review Group, which will review all funding requests.

Meanwhile, the expected departure of Senator Arlen Specter (Republican, Pennsylvania), the most active supporter of stem-cell research in the US Congress, from the chairmanship of the appropriations panel overseeing the NIH budget, could further weaken the chances of obtaining federal funds. ■

New regime may aim to craft compromise on green issues

Tony Reichhardt, Washington

Environmental scientists are waiting for signs of the new US administration's attitude towards basic environmental research. Areas such as global warming, air particulates, genetically modified food and endangered species have all been political battlegrounds in the past.

Although he strongly opposes the Kyoto climate treaty, during the presidential campaign George W. Bush came out in favour of research into the causes and impact of global warming. Otherwise, neither candidate mentioned the environment much.

As president, Bush's top environmental priority will be to devolve federal pollution controls. He views these as a matter for individual states and industry, rather than the government in Washington, to determine.

Bush's preference for the local, voluntary approach alarms the environmental lobby, which recalls attempts by President Ronald Reagan's administration to roll back federal environmental regulation in the 1980s. Environmentalists hope that Bush will not be as heavy-handed as Reagan, who hired controversial figures such as Interior Secretary James Watt to dismantle programmes seen as unfriendly to corporations or land developers.

So far, Bush has taken advice from a dozen or so conservative policy experts and state office-holders. Many of these are skilled at crafting compromises between environmentalists and business, and some of them are likely to fill key jobs in the new administration.

Led by Christopher DeMuth, head of the American Enterprise Institute, a conservative think-tank, they began advising Bush in mid-1999 on issues ranging from land use to global warming. This informal group of laissez-faire environmentalists includes Gayle Norton, former attorney general of Colorado; Lynn Scarlett of the Los Angeles-based Reason Public Policy Institute; Robert Nelson of the Competitive Enterprise Insti-

tute; Montana Governor Marc Racicot; and Richard Schmalensee, dean of the Sloan School of Management at the Massachusetts Institute of Technology, who has advised Bush on the economics of climate treaties.

One member of the group touted as a possible administrator of the Environmental Protection Agency (EPA) is Mary Gade, who spent 13 years working for the agency and is currently head of Illinois' environmental agency. Bush is also said to be considering two other heads of state environmental agencies, James Seif of Pennsylvania and David Struhs of Florida, for the post.

Racicot has long been seen as a potential secretary at the Department of the Interior, which administers the Endangered Species Act and includes the US Geological Survey. Another contender is Slade Gorton, who has just lost his Senate seat in the state of Washington, and who used to chair the Senate subcommittee overseeing the department's budget.

Such appointments will point the way for Bush's environmental policy, says Peter Saundry, executive director of the National Council for Science and the Environment, which lobbies for environmental research. Unlike Al Gore, who as vice-president was deeply involved in environmental policy, Bush will be inclined to delegate, Saundry says. This will give agency heads and their subordinates more leeway to shape policy.

Funding for research at the EPA, and at the interior and agriculture departments will be another key measure of the administration's commitment to environmental protection. The brightest prospects for environmental science may be at the National Science Foundation, which has been planning a dramatic expansion of its environmental research programme. Saundry is hopeful that, given past Republican calls for sound science to underpin environmental policy, this expansion will proceed under the new administration. ■



On a green path? Bush discusses environmental policies with landowners in Washington state.



Back on course: NASA has revived its plans to send a probe in 2004 to Pluto, shown here on the bottom left with its moon Charon. The decision could delay any mission to Jupiter's moon Europa until 2010.

NASA set to move Pluto back up its priority list

Tony Reichhardt, Washington

NASA hopes to announce a revival of its plans to send a spacecraft to Pluto in 2004. The turnaround may be confirmed as early as this week.

In a rapid reversal of its September decision to cancel the Pluto mission, the space agency may also allow teams outside its own space research centres to submit plans to design and build the spacecraft.

Ed Weiler, NASA's science chief, halted work on the Pluto-Kuiper Express after the combined cost for it and an associated mission to orbit Jupiter's moon Europa grew from \$654 million to \$1.5 billion (see *Nature* 407, 933 and 408, 505; 2000).

That decision, along with other cancelled missions, set alarm bells ringing in the planetary science community. It also forced scientists who advise NASA to state their priorities more clearly.

Michael Drake of the University of Arizona, representing NASA's advisory subcommittee for Solar System exploration, wrote on 27 November to the head of the agency's planetary programme: "It is logical to consider flying [the Pluto mission] first."

Until the Drake letter, Weiler says, planetary scientists had not expressed a clear preference over whether the Pluto or Europa mission should launch first if NASA had to make a choice. "That's the big change," he says.

Weiler has other reasons for sending the Pluto mission back to the head of the queue. Its launch is the more time dependent of the two projects: after December 2004, there will not be another favourable alignment of the planets for a decade, whereas the opportunity to visit Europa comes every 13 months.

Perhaps more importantly, placing an orbiter around Europa has also proved to be an extremely difficult engineering problem, owing to the severe radiation environment at Jupiter and the large amount of energy required to slow a spacecraft into orbit around the moon.

Given these challenges and current budgetary constraints, Weiler says: "If you go to Europa first, you're not launching until 2007 or 2008."

Many scientists outside NASA blame the spiralling costs on the coupling of the Pluto and Europa missions and a NASA directive that the projects must use the same technology. They contend that developing them as separate missions should slash the Pluto mission's price tag.

NASA originally handed both projects to its Jet Propulsion Laboratory (JPL) in Pasadena, without inviting outside proposals. But now, says Weiler: "It's entirely possible that there could be an [announcement of opportunity] in January for a competition to propose a Pluto mission."

Lockheed Martin Astronautics in Denver and the Applied Physics Laboratory at Johns Hopkins University in Baltimore, which operates the current Near Earth Asteroid Rendezvous spacecraft and has taken the lead in several NASA planetary missions, have both been working on ideas for trips to Pluto, and would probably answer any call for proposals.

Drake's letter pointed out that even if a mission to Europa was launched second, it would still return data before the Pluto spacecraft, because of the shorter travel time to Jupiter. But, says Weiler, this turns out to be not quite true, and "is only doable with [additional] money that we don't have right now."

If the Pluto spacecraft goes in 2004, NASA could not launch the Europa mission until 2010, given the JPL's current plan and cost estimates.

Data would then not come back until 2012 or 2013, compared with 2012 from Pluto. But Weiler hopes that, given more time to develop the technology, the costs of a Europa orbiter could fall.

He expects to make his decision before the end of 2000, and says: "If there's any chance to do Pluto within the available budget, I'm going to find a way to do it."

Japanese genomics company offers shares in sequences

David Cyranoski, Tokyo

A Japanese genomics company has developed a business model in which investors can buy shares in any of the genome sequences the company deciphers.

The idea comes from Takara Shuzo, a brewery company that plans to turn its subsidiary, Dragon Genomics, into Asia's biggest gene-sequencing operation (see *Nature* 404, 913; 2000).

As sequencing becomes faster and cheaper, the growing number of projects require new ways of financing that do not rely on company or government funding. The investment model for Dragon will be designed to attract investors who are interested in a particular project.

Sequencing projects are planned for the chimpanzee, silkworm, tuna and whale genomes, as well as for certain seaweed and mushroom species thought to have medicinal value.

Shares in the projects should become available during 2001, and profits from licensing agreements following on from the projects' results will be split among investors.

Companies such as the computer-games maker Konami have used an investment model based on the funding of individual projects in the past. But Takara believes it is the first company to do this for biotechnology.

It has applied for a 'business method patent' in Japan to cover its model, and is considering similar applications in other countries. "Biotechnology is a very competitive field," explains a Takara representative, "and this is an era for patenting everything and anything."

Takara will set up a separate company to handle the distribution of information to pharmaceutical companies and other interested parties, and to distribute profits to investors. The company will run a web page where investors will be able to monitor the latest status of sequencing, patents and licensing agreements.

Dragon plans to start operations next spring, and to have 45 state-of-the-art sequencing machines up and running by the summer of 2002. Most of its work will be done on a contract basis for corporations and academic researchers that have "money, but little technological capacity", says Takara's representative.

Italian women meet glass ceiling in the lab

Alison Abbott

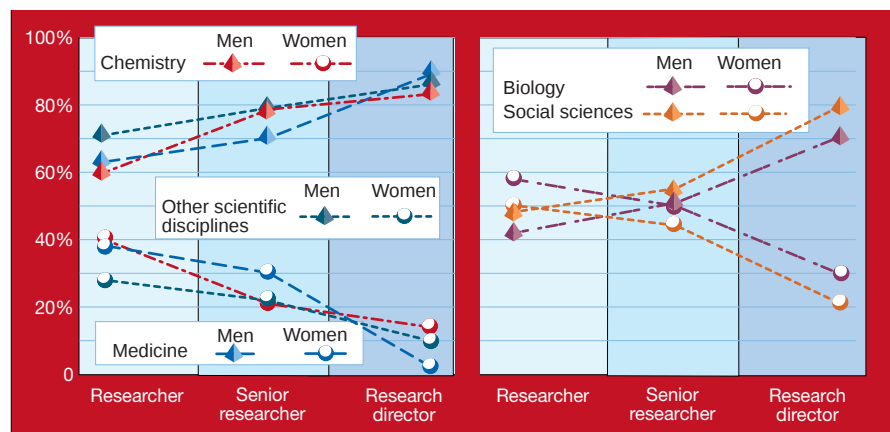
Men are three times more likely than women to be promoted to top positions in Italian publicly funded research laboratories, according to a study to be released early in the new year.

The study also finds that 90% of institute directors are male. Men even hold many of the top positions in disciplines where women form the majority of junior researchers, such as biology and the social sciences.

The study was conducted by a working group of senior women scientists, with the backing of Italy's National Research Council (CNR). It was chaired by Rossella Palomba, director of the CNR Institute for Population Research in Rome. Data were analysed from 15,000 publications and from the archives of hundreds of public research institutes, employing 7,000 scientists.

Palomba says that Italy is probably not the worst performer in Europe, but that the study's findings are still shocking.

"The difference in career opportunities for women in research in Italy cannot be explained by the difference in the number of publications," she says. On average, men publish under one paper per year more



Gender issues: how women's careers take a turn for the worse.

than women, she explains. But a bigger publication gap comes between the ages of 35 and 39, at a time when women are often involved in child care and men are busy getting promoted.

Many of the more successful women researchers forgo the opportunity to have children, the study finds. Nearly 40% of female research directors are childless, as are 22% of senior researchers.

Palomba suggests that the situation

might be improved if more women sat on selection panels, and if institutes were asked to publish the number of women on such panels. "Merely publishing gender statistics will probably make a difference," she says. "And having women on the panels will help break up the old-boy networks that cause women to be overlooked."

But Enzo Iarocci, president of the National Institute of Nuclear Physics (INFN), says that there are so few women in

top positions that gender correction on committees is not feasible.

"I was shocked to realize that there is a factor-of-ten loss between women who join INFN at the lowest rank and those at the top," he says. He sees no obvious solution, but says the INFN is "open to all suggestions".

One concept that finds no favour is the introduction of quotas of senior slots for women scientists. "We don't want a quota for women," says Flavia Zucco, a senior researcher at the CNR Institute for Biotechnology Research in Rome. "We just want the quota for men abolished."

The findings were welcomed by European Research Commissioner Philippe Busquin, who said the study was "the first to undertake a full comparative analysis of women's careers in public research institutes". Lack of data on women's careers in research hinders attempts to resolve the problem, according to Busquin.

Mary Osborn, a researcher at the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany, who helped produce the European Commission's recent report on women and science (see *Nature* 402, 337; 1999), sees "no single solution to the problem". She argues that the issue needs to be "tackled by all means and at all levels".

Another investigation of women's careers in Italian universities is about to start. ■

International search underway in Italy for institute directors

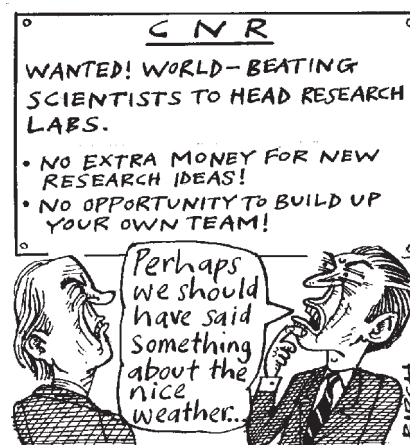
Italy's National Research Council (CNR) is this week breaking from its much-criticized practice of filling key posts from within its ranks, and is advertising internationally for scientists to head some of its institutes.

The positions are the first to arise under a reorganization that will combine some 300 CNR institutes and university-based centres into 95 larger units. This is the first step in a reform programme to make the CNR more efficient (see *Nature* 394, 712; 1998).

The appointments are for four-year renewable terms. Each new director will be responsible for coordinating research programmes at the merged centres and institutes.

Most scientists agree that the new arrangement will promote collaborative research, but say that, with no extra money available for research, it is not ideal.

Some observers are also sceptical about whether outsiders will find the new directorships attractive. For example, they will have few opportunities to build their own research teams. This is "a weakness in the process", says John Guardiola, head of



the CNR International Institute of Genetics and Biophysics in Naples.

Arturo Falaschi, the director of the International Centre for Genetic Engineering and Biotechnology in Trieste, is more optimistic. Together with an expected competition to attract 1,000 researchers, he says advertising the directorships could help to address problems such as gender imbalance.

A.A.

Big budget rise keeps NIH on course to double in five years

Washington The US Congress and President Bill Clinton's administration have agreed a budget for the National Institutes of Health (NIH) that will give the research agency a \$2.5 billion increase in funding for the 2001 fiscal year, which began on 1 October.

The 14% increase is the third substantial increase for the agency in as many years, and keeps the NIH on course to double its budget over five years.

Researchers were beginning to worry that the expected rise would not materialize (see *Nature* **408**, 627; 2000), but the agreed increase is only marginally below the \$2.7 billion increase Congress had supported.

Telecoms mogul pays for photonics research centres

San Diego Two research centres specializing in photonics are to be established at Stanford University in California and Duke University,

North Carolina, using a \$50 million gift from a telecommunications executive.

Each university has received \$25 million from Michael Fitzpatrick, a former chief executive of E-TEK Dynamics, which makes parts for fibre-optic networks. The centres will look at the use of light in telecommunications, and will emphasize

Singled out: \$50m bequeathed to fibre-optics research.

research collaboration with photonics-related corporations, university officials say.

Russia and China sign materials science deal

Tokyo The Russian Academy of Sciences and the Chinese Academy of Sciences have signed a memorandum agreeing to cooperate in materials science research.

Under the agreement, nanotechnology research centres will conduct joint projects, facilitate scientific exchange and promote the industrialization of their results. Leaders of the academies plan to meet regularly to ensure that the agreement is being followed.

CERN council pulls plug on its Higgs-hunting collider

Geneva Last week the council of CERN, the European Laboratory for Particle Physics, approved the dismantling of the laboratory's

Large Electron-Positron (LEP) collider. During the summer the LEP, almost at the end of its planned lifespan, glimpsed signs of the Higgs boson, a fundamental particle. This prompted many physicists to call for an extension to the collider's life.

Construction of the LEP's more powerful successor, the Large Hadron Collider, will start in the spring. With seasonal joviality, CERN's president, Luciano Maiani, proclaimed: "*Le roi est mort. Vive le roi!*"

Million-dollar prize for Canadian chemist

Montreal Howard Alper, a chemist at the University of Ottawa, has been named as the first winner of a Can\$1 million (US\$660,000) prize for achievement in Canadian science.

The Gerhard Herzberg Canada Gold Medal for Science and Engineering has been awarded to Alper for his work on synthesizing and modifying molecules.

The award, made by the Natural Sciences and Engineering Research Council and named after the 1971 Nobel prizewinner in chemistry, will give Alper money for research over the next five years, in addition to his existing grant support.

US government harmonizes gene-therapy regulations

Washington The National Institutes of Health (NIH) has proposed altering its requirements for the reporting of adverse events during clinical gene-therapy trials to match those of the US Food and Drug Administration (FDA).

Under the proposals, scientists would have 15 days to report unexpected reactions to an experimental treatment. Currently, they are supposed to report such reactions immediately — but do not always do so. NIH officials say the change should improve compliance with the rules.

Scientists had complained that the separate reporting requirements at the NIH and FDA were an unnecessary burden. Unlike the FDA, the NIH makes public all reports of such events that it receives.

Ireland upgrades Internet connections

Dublin Ireland is seeking to reinforce its position as a high-technology economy by boosting the speed of its researchers' Internet connections with the United States by a factor of 20.

The move will connect Irish research institutions to 180 US universities that are already participating in the high-speed Internet2 and Next Generation Internet projects. Ireland's participation was announced last week by Bertie Ahern,



Hands-on: Ahern (left) and Clinton last week.

the prime minister, during a visit by US president Bill Clinton.

The change will underpin Ireland's IR£1.95 billion education and research plan, which was announced earlier this year.

Japan brings in measures to prevent spread of BSE

Tokyo Japan's ministry of health has taken steps to stop the circulation of products made with "dangerous" parts of cattle, sheep and goats. It is responding to new concerns about the spread of bovine spongiform encephalopathy (BSE) and cases of the human form of the disease, variant Creutzfeldt-Jakob disease, in Europe.

The ban requires makers of cosmetics, medical devices and pharmaceuticals to inspect any products made using brains, eyes or intestines of animals from 29 countries, mostly in Europe.

The government says this is a preventive measure. There have been no known cases of variant Creutzfeldt-Jakob disease in Japan.

Michigan's tobacco payouts go to science

Washington The state of Michigan has launched a billion-dollar effort to build up its life-sciences research. The programme will be funded by tobacco corporations under a legal settlement (see *Nature* **400**, 391; 1999).

Michigan has 20 years to spend the money; some \$90 million in projects has been allocated so far. The University of Michigan in Ann Arbor will get \$48 million, with \$12 million for a proteomics facility and \$9 million to set up a bioinformatics centre. Michigan State University in East Lansing gets \$26 million to set up a structural-biology centre. Wayne State University in Detroit and the Van Andel Institute in Grand Rapids are also involved in the state's plan.

Correction The recent Opinion article "Rays of hope in eastern Europe" (*Nature* **408**, 625; 2000) should have included an acknowledgement of the European Science Foundation as a co-sponsor of *Nature's* Dresden meeting on that topic.

2000 in context

It was the year of genomes. Every week seemed to bring another landmark — be it human, animal, plant or pathogen. But there was more to 2000 than strings of 'A's, 'C's, 'G's and 'T's. *Nature* explores some of the highs, lows and emerging trends behind the year's scientific headlines.

Genomics

Beyond the book of life

Where were you on 26 June 2000, while history was being made? On that day, at press conferences throughout the world, scientific leaders ceremonially opened the 'book of life', announcing the completion of a working draft of the human genome.

In Washington, leaders of the publicly funded Human Genome Project (HGP) made peace at the White House with their bitter rival, Craig Venter's Celera Genomics of Rockville, Maryland — which simultaneously declared complete its own 'first assembly' of the human genome. President Bill Clinton, the consummate politician, provided the necessary rhetoric: "With this profound new knowledge, humankind is on the verge of gaining immense new power."

In reality, the chosen date was arbitrary. But by neatly bisecting 2000, the joint HGP/Celera announcement provided an appropriate point to mark the transition from the genome sequencing era to the age of functional genomics. Although much DNA remains to be sequenced, the emphasis will now be on deciphering what the accumulated sequence information means. And that is where the scientific fun really starts.

"The stage is set for a full scale exploration of the ways in which this disarmingly simple one-dimensional instruction book is converted into the four dimensions of space and time that characterize living organisms," says Francis Collins, director of the National Human Genome Research Institute in Bethesda, Maryland.

Despite its big budgets and high public profile, genomics has so far been seen by many biologists as a subdiscipline of genet-

ics. But with the expansion of functional genomics, it will soon be influencing every area of biology. And given concerns about the implications of gene patenting, and privacy of genetic information, genomics also promises to keep social scientists, ethicists and lawyers extremely busy.

In the first half of this year, relations between the HGP and Celera became seriously strained, as the rivals failed to agree terms for collaboration. The impasse was the issue of access to the sequence data. The HGP refused to accept restrictions demanded by Celera, which argued that it had to protect its commercial interests. But in the end, political pressure on the two sides to stop the squabbling — which was in danger of undermining public recognition of both groups' achievements — proved sufficient to broker the joint announcement, if not a meaningful collaboration.

The quality of the HGP's draft and Celera's assembly still remains unclear. Peer-reviewed papers providing a detailed synthesis of each group's work are expected to appear early in the New Year. Filling in the remaining gaps in the sequence, and repeated sequencing by the HGP to correct any errors in the code, could take until 2003. But one chromosome, number 21, was unveiled in finished form in May¹ — the second to be completed after chromosome 22 (ref. 2).

This year's model

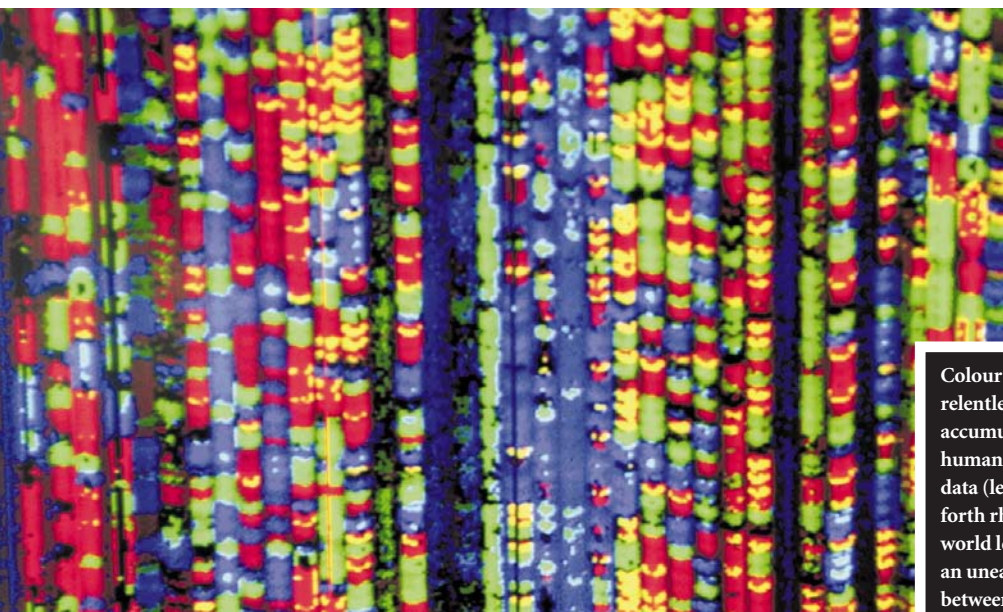
While we wait for publication of the human genome, there have been plenty of other sequencing milestones this year. Laboratory 'model' organisms led the way, with the fruitfly *Drosophila melanogaster*³ and the

workhorse of plant science, *Arabidopsis thaliana*⁴, being the published highlights. Complete sequences for pathogenic microorganisms also came thick and fast, including the cholera bacterium *Vibrio cholerae*⁵ and *Pseudomonas aeruginosa*⁶, a common cause of opportunistic infections. And among the pathogens came a paper proving that genomics is not the preserve of developed nations. In July, a Brazilian consortium completed the genome of *Xylella fastidiosa*⁷, which causes variegated chlorosis, a disease of citrus crops.

As each organism's sequence is completed, the focus shifts to characterizing all of its genes, and determining their functions — or 'annotating' the genome. For *D. melanogaster*, this task was kicked off in late 1999 with a two-week jamboree held at Celera's Rockville headquarters, at which the company's scientists worked alongside academic fruitfly biologists. Annotating the human genome may require innovative 'collaboratory' strategies, with scientists sharing data over the Internet.

The first job is to determine the total number of human genes, now generally thought to lie somewhere between 30,000 and 70,000. Geneticists and bioinformaticists are running a sweepstake on the outcome — see <http://www.ensembl.org/Genesweep> for the current distribution of bets. But going beyond gene numbers to investigate gene function is where things really start to get difficult.

Comparisons between species can help. Hints at the function of unknown genes may come from similarities to sequences in well-studied genes from model organisms. More generally, cross-species comparisons can identify conserved sequences involved in processes of fundamental biological significance, and help understand the genome's overall structure.



Colour coded: the relentless accumulation of human genome data (left) brought forth rhetoric from world leaders and an uneasy truce between Celera head Craig Venter and the publicly funded effort.



BOB BOSTON, WASHINGTON UNIVERSITY

This is why fruitfly biologists are gearing up to sequence a second *Drosophila* species, probably *D. pseudoobscura*, and researchers annotating the genome of the nematode worm *Caenorhabditis elegans* would like to sequence *C. briggsiae*. For the human genome, a complete mouse genome sequence will be similarly valuable — and Celera and publicly funded researchers are working separately towards this new goal.

Although much progress can be made using computational tools that compare gene sequences, or predict protein structure from DNA sequence information, the mouse sequence will also offer another advantage to those annotating the human genome. It is possible to experiment with mice and other model organisms in a way that is impossible with people, disabling genes systematically to determine their function.

Indeed, functional genomics will require a diversity of techniques to disrupt normal gene activity. Although it probably cannot be generally applied in mammals, one of the most promising is a phenomenon called RNA interference. This is the gene-silencing effect that occurs when cells are exposed to double-stranded RNA matching the sequence of a given gene. The technique has already been applied systematically in *C. elegans*: two papers published in November^{8,9} targeted two of the worm's six chromosomes, identifying many genes involved in development and cell division.

Some researchers are using small molecules to disable particular genes and so determine their precise functions. This field of 'chemical genetics' scored a notable success in September. By applying some subtle genetic manipulation to render genes susceptible to a particular chemical 'switch', a team of chemists and geneticists showed that the approach can be used to selectively disable any protein kinase¹⁰. This large family of

enzymes has previously proved particularly resistant to detailed functional analysis.

Other research groups are using chemical genetics in a less directed way, adding a range of small molecules to cells until they find one that causes a particular effect, and then working out which gene the molecule has disrupted¹¹. This approach is conceptually similar to 'forward genetic' screens that randomly mutate the genome of model organisms to create a wide range of biologically interesting mutants. Two such screens for mutant mice, unveiled in August^{12,13}, are expected to become important resources for those interpreting the mouse genome.

Code breakers

But ultimately, no amount of experimental studies on model organisms will uncover all of the secrets hidden in the human genetic code. And on the human side, one of the top priorities is to identify the subtle genetic variations that make people susceptible to big killers such as cancer and heart disease. To this end, several large population studies examining genetic variation, lifestyle factors and disease are now getting underway throughout the world.

This task has been helped by the efforts of a pioneering collaboration between academic groups, multinational companies and Britain's charitable Wellcome Trust. The Single Nucleotide Polymorphism Consortium is preparing a map of many hundreds of thousands of genetic markers that will be an invaluable resource for anyone trying to pin down the location of disease genes (see <http://snp.cshl.org>).

Other researchers believe that the key to turning the human genome into tomorrow's drugs will be the industrialization of structural biology. The leaders of this new field of structural genomics¹⁴ are now busy creating

'protein structure factories'. Here, DNA sequences will be engineered into cells to culture large quantities

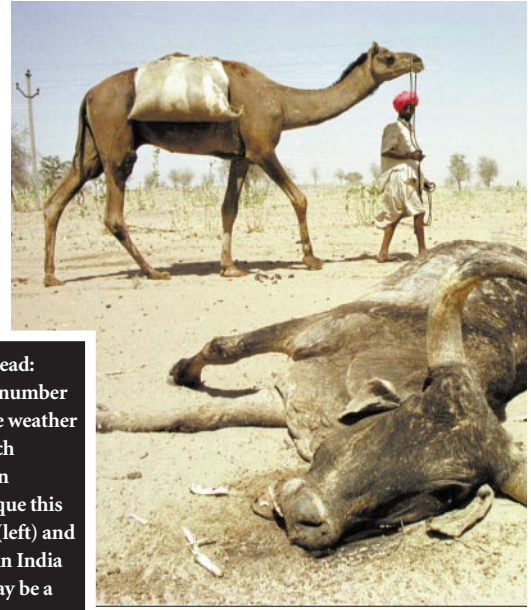
of protein, which will be purified and subjected to structural analysis. The goal is to automate the entire process, churning out protein structures on an unprecedented scale.

But even these huge efforts are still based on analysing genes and proteins one by one. In real biological systems, genes and proteins work in concert — which is why perhaps the biggest growth area within functional genomics in 2000 was the use of DNA microarrays, or 'gene chips', to investigate wider patterns of gene expression.

A typical microarray consists of hundreds, or even thousands, of DNA sequences from the coding regions of individual genes immobilized on a surface. Messenger RNA from a given sample will bind to the corresponding sequence on the chip. As a result, DNA microarrays allow biologists to see at a glance which genes are active within a given cell, or tissue. Among this year's highlights were two studies of the yeast *Saccharomyces cerevisiae* — one examining patterns of gene expression associated with particular signal transduction pathways¹⁵, the second documenting changes in gene expression caused by mutations and exposure to a range of chemicals¹⁶. Microarrays were also used this year to profile gene expression in human cancers^{17,18}.

Microarray technology, this time involving immobilized proteins¹⁹, could prove crucial in what is the biggest functional genomic challenge: proteomics. This field aims to understand the function of every protein produced by an organism — an enormous task, given that processes such as RNA editing may allow the tens of thousands of genes within the human genome to produce several million distinct proteins. Proteomics received a boost in February, with a pioneering study of protein-protein interactions in *S. cerevisiae*²⁰. But

LIAMSON AGENCY



Storms ahead: the rising number of extreme weather events, such as floods in Mozambique this February (left) and droughts in India (right), may be a sign of climate change due to human activities.

AP

in the long run, whether or not the discipline achieves its ambitious goals may depend on the development of advanced new technologies for protein analysis.

Companies and academic groups worldwide are now jostling for position on this wild functional genomic frontier. And true to form, Venter vows that Celera will be among the leaders. If that provides the same competitive spur to this field as it did in the race to sequence the human genome, prepare for another fast-paced year.

Peter Aldhous

1. The Chromosome 21 Mapping and Sequencing Consortium *Nature* **405**, 311–319 (2000).
2. Dunham, I. *et al.* *Nature* **402**, 489–495 (1999).

3. Adams, M. D. *et al.* *Science* **287**, 2185–2195 (2000).
4. The *Arabidopsis* Genome Initiative *Nature* **408**, 796–815 (2000).
5. Heidelberg, J. F. *et al.* *Nature* **406**, 477–483 (2000).
6. Stover, C. K. *et al.* *Nature* **406**, 959–964 (2000).
7. Simpson, A. *et al.* *Nature* **406**, 151–157 (2000).
8. Fraser, A. G. *et al.* *Nature* **408**, 325–330 (2000).
9. Gönczy, P. *et al.* *Nature* **408**, 331–336 (2000).
10. Bishop, A. C. *et al.* *Nature* **407**, 395–401 (2000).
11. Mayer, T. U. *et al.* *Science* **286**, 971–974 (2000).
12. Nolen, P. M. *et al.* *Nature Genet.* **25**, 440–443 (2000).
13. Hrabé de Angelis, M. *et al.* *Nature Genet.* **25**, 444–447 (2000).
14. *Nature Struct. Biol.* **7**, Suppl., 927–994 (2000).
15. Roberts, C. J. *et al.* *Science* **287**, 873–880 (2000).
16. Hughes, T. R. *et al.* *Cell* **102**, 109–126 (2000).
17. Alizadeh, A. A. *et al.* *Nature* **403**, 503–511 (2000).
18. Perou, C. M. *et al.* *Nature* **406**, 747–752 (2000).
19. MacBeath, G. & Schreiber, S. L. *Science* **289**, 1760–1763 (2000).
20. Uetz, P. *et al.* *Nature* **403**, 623–627 (2000).

Global warming

A climate of uncertainty

If so much was not at stake, the circumstances under which last month's climate talks collapsed might raise an ironic smile. Tired negotiators, working into the early hours of the morning, ran out of time. If they had been able to meet the next day, maybe they could have struck a deal. But the Netherlands Conference Centre in The Hague had been booked for a meeting of oil industry executives — the very people whose products are warming the world.

The talks' failure leaves the Kyoto Protocol, the international treaty to limit greenhouse gas emissions, in limbo. Three years ago in Japan, developed countries agreed to cut their carbon dioxide emissions by an average of 5% of their 1990 levels by 2010. But the details of how the cuts would be made remained to be thrashed out. And without this fine print, not one major industrialized nation has been prepared to

ratify the protocol.

The meeting in The Hague was supposed to accelerate progress. But the position of the United States was always likely to be a stumbling block. US negotiators went into the meeting determined that their nation should be allowed to offset against its emission targets changes in the management of its forests and farmland to increase their potential as carbon sinks. Sinks are ecosystems that pull carbon dioxide from the atmosphere and convert it into biomass. Similarly, the United States also wanted 'carbon credits' for planting forests in developing countries.

By the end of the meeting, the last of these demands had been dropped. US negotiators had also agreed to put a ceiling on their domestic sinks of 75 million tonnes of carbon a year — a quarter of their original request. But the European Union

still found this unacceptable.

What happens next is unclear, with the position of US President

Elect George W. Bush being crucial. If his administration were to strike a deal with Europe to limit carbon dioxide emissions, Bush will probably have more chance of pushing it through a sceptical Senate than his defeated opponent Al Gore would have had. But Bush's close links with the oil industry make environmentalists pessimistic about his will to act.

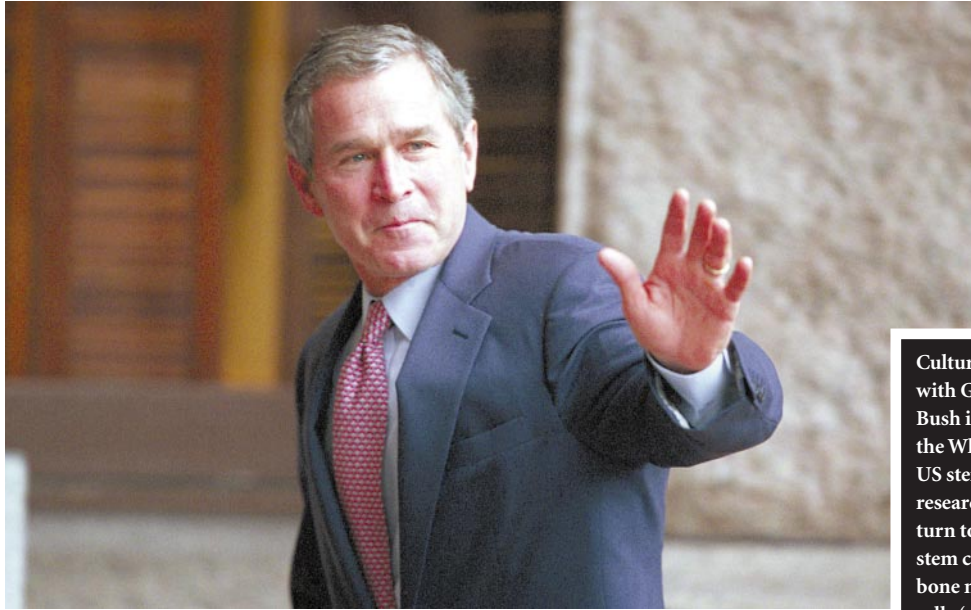
Some analysts remain hopeful that a deal will be struck in the spring. Others argue that Europe, Japan and Russia should push ahead without the United States. Yet others say that the Kyoto Protocol is a flawed treaty that was never going to address climate change seriously. They would be happy to tear it up and start again.

But ultimately, some sort of international agreement is needed, and key to its success will be a better understanding of the implications of rising atmospheric concentrations of greenhouse gases, and of the behaviour of carbon sinks.

This year has seen a strengthening of the consensus that global warming is real. Sceptics had pointed previously to discrepancies between temperature measurements made at the surface and from satellites. But in January, a report from the US National Academy of Sciences concluded that the data were now reconciled, and that they pointed to a warming planet¹.

Elsewhere, however, more knowledge has only increased uncertainty. For instance, atmospheric aerosols of dust, soot and other particles have long been thought to exert a net cooling effect, countering the build-up of greenhouse gases by blocking out sunlight. But in May, two papers from the international Indian Ocean Experiment revealed that soot can turn the net effect of

AMIT BHARGAVA/NEWSMAKERS



Culture shock: with George W. Bush installed in the White House, US stem-cell researchers may turn to using adult stem cells, such as bone marrow stem cells (right), rather than embryonic cultures.



JOE READ/NEWMAKERS

EYE OF SCIENCE/SPL

aerosols into warming — directly, by absorbing solar radiation², and indirectly, by reducing cloud cover³.

The future behaviour of carbon sinks is highly uncertain. Last month, a simulation by researchers at Britain's Meteorological Office, using a model of the flow of carbon between the ocean, the atmosphere and the biosphere, suggested that terrestrial sinks may become net producers of carbon dioxide within 50 years⁴. Given uncertainties in current models of the carbon cycle, this is unlikely to be the final word on the matter. And the difficulty of assessing the effects of carbon sinks was underlined by papers revealing that the amount of carbon stored shows large fluctuations in space and time^{5,6}.

The impacts of forests on global warming are further complicated by factors other than their role as carbon sinks. At high latitudes, planting forests can replace open, snow-covered ground with a dark canopy. The result is that less sunlight is reflected in winter, causing warming that could outweigh the effects of carbon uptake⁷.

Results from palaeoclimatology confirm the picture still further. Earlier this month, an analysis of climate and atmospheric carbon dioxide over the past 540 million years questioned the role of carbon dioxide as the main driver of climate change over geological timescales⁸. But this does not mean that we can be complacent about the much more rapid build-up of greenhouse gases resulting from human activities.

Clearly, climate researchers have a lot to do. One of their predictions is that global warming will increase the incidence of extreme weather events such as droughts, storms and floods⁹. Although it is impossible to blame global warming for individual natural disasters — such as the floods that devastated Mozambique in

February and March — they do seem to be more frequent. Even the British, used to treating their gentle but dreary climate as a topic for polite conversation, have over the past few months been hit by the heaviest floods on record. If such trends continue, the pressure on climate scientists to provide politicians and the public with more certainty will intensify.

Heike Langenberg and Peter Aldhus

Stem cells

Panacea, or Pandora's box?

Biology is continually throwing up ethical dilemmas. And the thorniest issue of 2000 was that surrounding research on human embryonic stem cells.

These cells can, in theory, give rise to any of the body's tissues — and so have enormous potential in regenerative medicine. But 'pro-life' activists are bitterly opposed to their use, as cultures of embryonic stem cells can only be created by destroying a pre-implantation human embryo.

The freedom of researchers to work with embryonic stem cells rests on a knife-edge in many countries. In the United States, home to the largest community of stem-cell researchers, the National Institutes of Health proposed in August that federal funds could be used to support the research, provided the cells had been isolated using private money. But with the election of George W. Bush to the presidency, this decision might be rescinded.

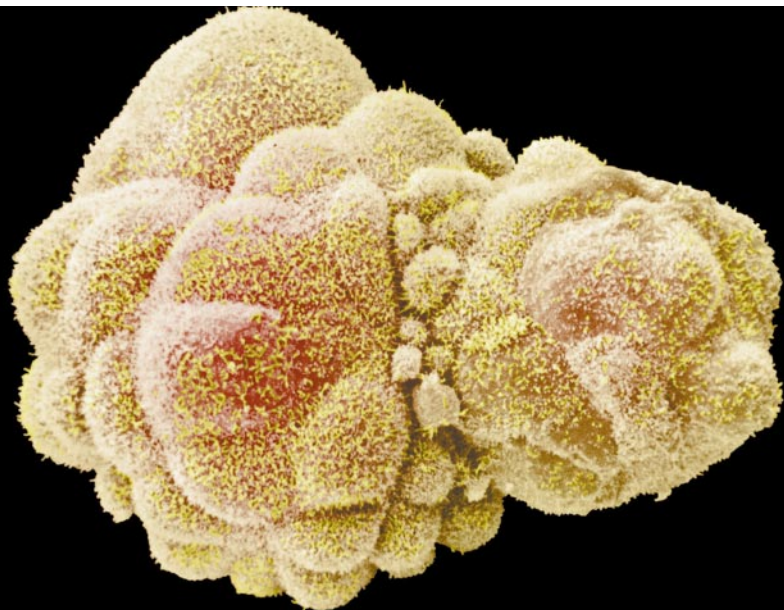
Campaigners against embryonic-stem-cell research have made much of experiments suggesting that adult stem cells, which reside in and help regenerate many tissues,

can give rise to a much wider variety of cell types than was thought. Given this remarkable plasticity, they argue, attempts to develop stem-cell therapies should not need human embryonic stem cells.

This unexpected phenomenon became well known in 1999, when a team led by Angelo Vescovi of the San Raffaele Hospital in Milan showed that neural stem cells taken from the brains of mice, cultured, and injected into bone marrow could give rise to apparently normal blood cells¹. This finding has proved difficult to replicate. But the past year has seen further demonstrations of plasticity, including two papers earlier this month that showed the reverse transformation — mouse bone-marrow stem cells giving rise to cells that resemble neurons^{2,3}. Vescovi's group encored in October by showing the mouse and human neural stem cells could produce muscle cells⁴.

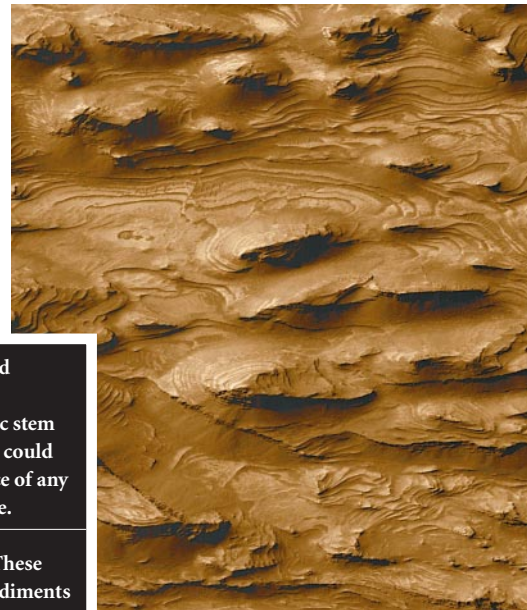
Although these experiments hint that adult stem cells have great therapeutic potential, most researchers in the field argue that this may never be realized if the door to research on human embryonic stem cells

1. <http://books.nap.edu/books/0309068916/html/index.html>
2. Satheesh, S. K. & Ramanathan, V. *Nature* **405**, 60–63 (2000).
3. Ackerman, A. S. *et al. Science* **288**, 1042–1047 (2000).
4. Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. J. *Nature* **408**, 184–187 (2000).
5. Battle, M. *et al. Science* **287**, 2467–2470 (2000).
6. Bousquet, P. *et al. Science* **290**, 1342–1346 (2000).
7. Betts, R. A. *Nature* **408**, 187–190 (2000).
8. Veizer, J., Godderis, Y., Francois, L. M. *Nature* **408**, 698–701 (2000).
9. Easterling, D. R. *et al. Science* **289**, 2068–2074 (2000).



Divide and conquer: embryonic stem cells (left) could be a source of any tissue type.

Red sea? These layered sediments on Mars may have been laid down in water.



is closed. The problem, they say, is that stem-cell science is in its infancy—its practitioners know little of the molecular mechanisms that underlie the plasticity of stem cells, or that cause them to travel down particular developmental pathways. For some tissues, they are still arguing over the precise identity of the stem cells involved. “We’re in the Dark Ages,” observes Leonard Zon, a developmental biologist at the Children’s Hospital in Boston.

Many researchers agree that, in the long term, adult stem cells may be preferred for therapeutic applications. But experiments with embryonic stem cells, which have the greatest developmental potential, will almost certainly be needed for these therapies to see the light of day. Stem-cell biologists reject the notion that research on embryonic stem cells from mice will be enough. John Gearhart of Johns Hopkins University in Baltimore points out that human and mouse embryonic stem cells behave differently in culture. “These cells do vary considerably,” he says.

Ron McKay, a stem-cell biologist at the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland, believes that the potential of adult stem cells will vary from disease to disease. “It depends absolutely on what you’re interested in,” he says.

For liver regeneration, work in mice suggests that adult stem cells have great clinical promise: in November, researchers with the company StemCells in Sunnyvale, California, cured animals of a lethal liver disease by transforming bone-marrow stem cells into liver cells⁵. Preliminary data from humans are also encouraging⁶.

But McKay’s work, on therapies for Parkinson’s disease that require the production of neurons that release the neurotransmitter dopamine, suggests that adult stem cells have their limits. McKay can generate

large numbers of dopamine-producing neurons from mouse embryonic stem cells⁷, but he has so far failed to achieve this using adult neural stem cells.

Researchers agree that the key to developing therapies is the fundamental understanding needed to drag the field out of its Dark Ages. And although this process is still in its early days, some papers published in 2000 gave chinks of light.

For instance, it is becoming clear that signals requiring cell-to-cell contact help determine stem cells’ eventual fate^{4,8}. Meanwhile, researchers are starting to discover the intracellular factors involved. In March, a team at the Massachusetts General Hospital showed that a protein called p21 is vital for maintaining bone-marrow stem cells — without it, the cells proliferate rapidly and are soon exhausted⁹. And in September, a Canadian-

German team showed that a transcription factor called Pax7 prevents the stem cells in muscle from developing into other tissues¹⁰.

Despite these advances, stem-cell researchers warn that there is a mountain still to climb. “To manipulate the variables is extremely difficult,” says Vescovi.

Peter Aldhous

1. Bjornson, C. R., Ritzke, R. L., Reynolds, B. A., Magli, M. C. & Vescovi, A. L. *Science* **283**, 534–537 (1999).
2. Brazelton, T. R., Rossi, F. M. V., Keshet, G. I. & Blau, H. M. *Science* **290**, 1775–1779 (2000).
3. Mezey, E., Chandross, K. J., Harta, G., Maki, R. A. & McKercher, S. R. *Science* **290**, 1779–1782 (2000).
4. Galli, R. *et al.* *Nature Neurosci.* **3**, 986–991 (2000).
5. Lagasse, E. *et al.* *Nature Med.* **6**, 1229–1234 (2000).
6. Theise, N. D. *et al.* *Hepatology* **32**, 11–16 (2000).
7. Lee, S.-H., Lumelsky, N., Studer, L., Auerbach, J. M. & McKay, R. D. *Nature Biotech.* **18**, 675–679 (2000).
8. Tsai, R. Y. L. & McKay, R. D. G. *J. Neurosci.* **20**, 3725–3735 (2000).
9. Cheng, T. *et al.* *Science* **287**, 1804–1808 (2000).
10. Seale, P. *et al.* *Cell* **102**, 777–786 (2000).

Space

Where are they boldly going?

NASA administrator Dan Goldin could hardly have started 2000 in a worse mood. After the embarrassing loss of two Mars missions in a row, the second just three weeks before Christmas, the space agency braced itself for a long winter of fact-finding reports, congressional hearings and what-went-wrong news stories.

What had gone wrong, it turned out, was the application of Goldin’s ‘faster, better, cheaper’ slogan, NASA’s guiding principle for much of the 1990s. Even before the last report landed on Goldin’s desk, the lesson was clear: his troops had been focusing on cheaper and faster at the expense of better.

A year later, the space agency —

particularly its space science division, which had embraced Goldin’s credo most enthusiastically — is still coping with the fallout. Projects started in the early 1990s were found, upon more realistic examination, to be overstretched and underfunded. And space scientists watched aghast as NASA science chief Ed Weiler recut his cloth accordingly.

Some problems have proved more tractable than others. Projects still in the planning stage, such as the Space Interferometry Mission to map stars and search for new planets, scheduled for launch in 2009, have been sent back to the drawing board to cut costs.



Rising star: after years of struggle, the International Space Station is now occupied, under construction and visible from Earth with the naked eye. It aims to provide a permanent laboratory in orbit.



AP

Others have had their schedules stretched. Realizing that the Jet Propulsion Laboratory (JPL) in Pasadena, California, might struggle to launch two major projects in the first quarter of 2001 — the Genesis spacecraft to investigate the solar wind in February and the Mars Odyssey orbiter in April — NASA bumped Genesis to June to allow some breathing room. And finalists in the competition to propose Small Explorer astronomy missions, who usually get money to hone their ideas, were told that they would have to wait a year for funding.

But in other cases Weiler resorted to flat-out cancellation. The Mars programme, most overextended of all, threw off ballast by scrapping a lander planned to accompany the 2001 Mars orbiter. Weiler also cancelled a tiny 'nanorover', built by JPL to fly on Japan's MUSES-C mission to visit an asteroid and return with samples in 2002; its budget had ballooned from \$20 million to \$60 million. And a long-proposed spacecraft to visit distant Pluto and the Kuiper Belt was shelved indefinitely to preserve the option of funding another ambitious mission to Jupiter's moon Europa.

The wave of cancellations has sparked a revolt among planetary scientists, and the Europa-versus-Pluto decision was challenged by advisory groups who say that NASA can still have both by taking some of the work away from the JPL and opening it to outside competition.

But painful as it has been, this self-examination may do NASA good. The revamped Mars programme, although less ambitious, is now generally seen as more realistic. NASA has backed away from a plan to send an orbiter and lander to the planet every two years, and has pushed a costly and technically challenging mission to collect and return with samples back a decade, to 2014 at the earliest. The plan now calls for

two modest-sized landers equipped with rovers in 2003, an orbiter carrying a very high-resolution camera in 2005, then a larger lander with a mobile laboratory in 2007. These spacecraft would be supplemented by smaller scouts dispatched to different areas of the planet's surface.

The new Mars roadmap is also flexible enough to accommodate new discoveries. This is especially important in the light of some of NASA's best news of the year — the results from the Mars Global Surveyor, which has been mapping the planet's surface. In two papers, Michael Malin and Kenneth Edgett of Malin Space Science Systems in San Diego reported new evidence of liquid water in the planet's history. Their camera revealed what may be the traces of springs in the sides of cliffs, which Malin and Edgett suggest may have flowed in the very recent past¹. Other photos show what appear to be sedimentary deposits, possibly the beds of ancient lakes, laid down between 3.5 billion and 4.3 billion years ago².

The first result is controversial, because scientists have great difficulty explaining how liquid water could exist near the surface of such a cold planet. "The reason for my scepticism is the physical conditions," says Michael Carr of the US Geological Survey in Menlo Park, California. But the second finding is considered the most compelling evidence yet for ancient Martian lakes, which have been hypothesized for years. Should robotic or human geologists land on the planet looking for fossils, they now have a better idea of where to search.

Scientists will not have to wait long for more data. A gamma-ray spectrometer onboard the Odyssey orbiter, due to arrive at the planet in October, should be able to detect hydrogen, and by inference water, within centimetres of the surface. And the

European Space Agency's Mars Express orbiter, set for launch in 2003, will carry a ground-penetrating radar capable of detecting water as deep as five kilometres below the surface. Any groundwater, says Carr, "should really shine like a beacon".

Another hopeful sign for NASA came in a field where it has been much criticized — biological and microgravity research done by astronauts in orbit. The agency has restructured its programme to play down experiments on the growth of protein crystals that have produced no conclusive results, and to focus on fundamental research in gravitational biology and physics. Buoyed by a few isolated but intriguing results from space shuttle experiments, where levels of gene expression in cultured cells have changed under weightless conditions³, NASA is looking forward to an era where astronauts can conduct ongoing laboratory research in orbit for the first time.

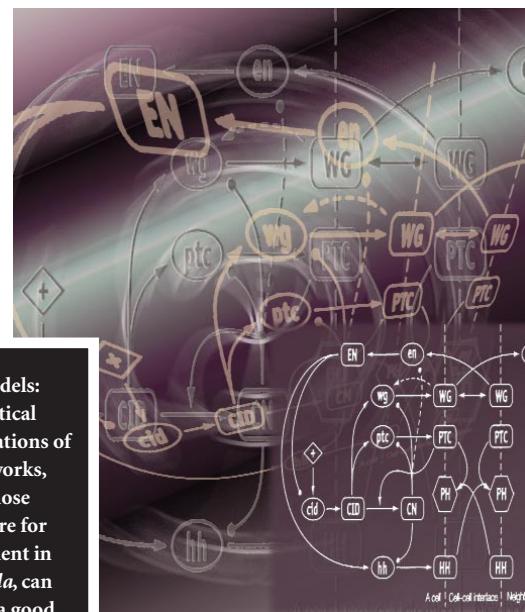
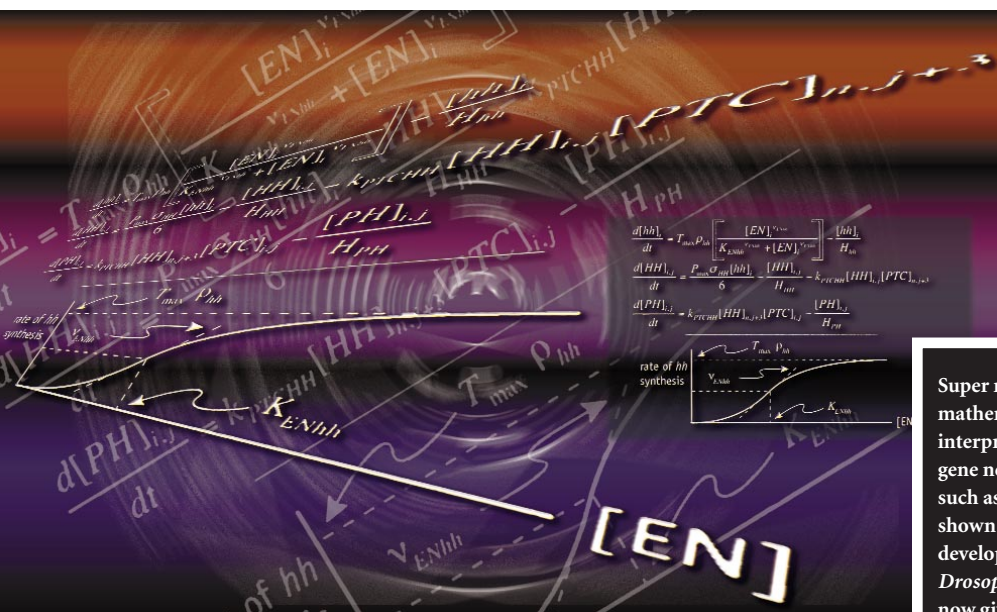
In fact, the International Space Station has been NASA's most pleasant surprise of the year. Although many scientists still regard it as a white elephant, it is the agency's flagship project. And after years of struggle to get the project underway, the station is now occupied and under construction. So far, the astronauts have made it look easy. And although there is plenty of scope for pitfalls, the project's managers could be forgiven a bit of optimism earlier this month as the station's main solar panels unfolded to create a new star in the heavens, just in time for Christmas and the dawn of 2001 — a date with symbolic importance for space aficionados everywhere.

Tony Reichhardt

1. Malin, M. C. & Edgett, K. S. *Science* **288**, 2330–2335 (2000).

2. Malin, M. C. & Edgett, K. S. *Science* **290**, 1927–1937 (2000).

3. Hammond, T. G. et al. *Nature Med.* **5**, 359 (1999).



Super models: mathematical interpretations of gene networks, such as those shown here for development in *Drosophila*, can now give a good description of real biological systems.

Mathematical biology

Life is a game of numbers

Gels, microscopes and pipettes are the stock in trade of the modern molecular biologist. But mathematical models of gene and protein networks could soon be just as important. And 2000 may go down as the year in which mathematics started to make an impact on mainstream biology.

"There's a lot of excitement and a lot of people moving into the field," says Dennis Bray of the University of Cambridge, one of its pioneers.

Although structural biologists and neuroscientists have long relied on mathematics to interpret their experiments, most molecular, cellular and developmental biologists have kept their distance. But with the accumulation of genomic data and the advent of techniques to monitor thousands of cellular components at once, things are going to change. "We're getting to the point of using models to make sensible predictions," says Bray.

The field's potential was hinted at in July by George von Dassow and his colleagues at the University of Washington in Seattle¹. Their aim was to simulate the behaviour of a group of fruitfly genes called the segment polarity network, which helps shape embryonic development, using a model made up of more than 100 differential equations.

But, try as they might, the researchers could not make their virtual genes behave like the fly's. After many weeks of tweaking parameters such as protein half-lives, diffusion constants and binding coefficients, the researchers looked at the components of their model again.

This revealed that it seemed to lack two key links. When von Dassow and his

colleagues trawled the literature, they uncovered a pair of studies that suggested two additional ways in which gene products could influence gene activity. The mathematical approach had revealed the significance of findings overlooked by most biologists. "In my opinion, the role of mathematical models is to tell you what you don't know," says Garret Odell, who heads the group in which von Dassow works.

Armed with this knowledge, von Dassow and his colleagues updated their model. They expected to have to fine-tune the activities of every gene and protein to make it work. But to their surprise, the model not only worked without a struggle, it could tolerate an enormous amount of slop. Roughly nine times out of ten, replacing one of the values in the model with a random number did not affect the gene network's overall function.

"This is a piece of engineering design that transcends all human ability," says Odell. "Almost everything humans make will fail if any single part gets a little out of tolerance or is a little bit wrong."

Stanislas Leibler and his colleagues at Princeton University in New Jersey, modelling bacterial movement in response to chemical signals, have found a similar breadth of tolerance². These findings suggest that such robustness may be a widespread characteristic of life, which has evolved to help cope with an unpredictable world.

Other researchers are starting to apply mathematical models to the manipulation of biological systems. For example, biomedical engineer James Collins at Boston University

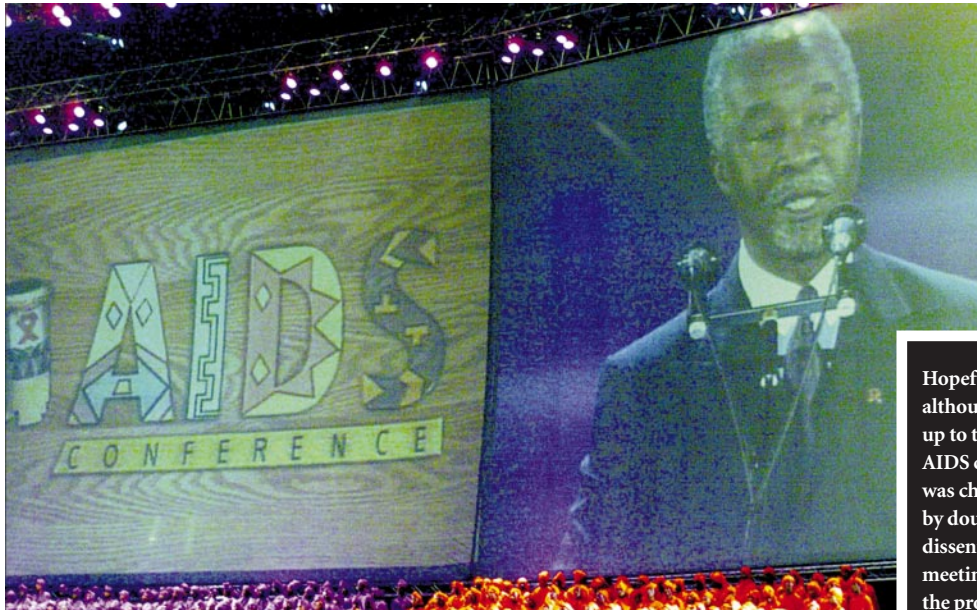
and his colleagues have used differential equations to design a circuit made up of a pair of genes that turn on and off in a mutually exclusive fashion in response to an external chemical signal — a genetic toggle-switch³.

They managed to genetically engineer the circuit into *Escherichia coli* bacteria. Working independently, Leibler's team has used an almost identical strategy to engineer a genetic oscillator into *E. coli* — a gene that switches on and off with a more or less regular period⁴.

But perhaps the surest sign of the growing importance of mathematical biology is the field's crop of new projects and even entire institutes. The eminent biologists Leroy Hood and Sydney Brenner, for example, have founded the Institute for Systems Biology in Seattle and the Molecular Sciences Institute in Berkeley, California, respectively.

Meanwhile, Nobel laureate Al Gilman at the University of Texas Southwestern Medical Center in Dallas has won a \$25 million, five-year grant for his Alliance for Cellular Signaling, which will rely heavily on mathematical modelling. The US National Science Foundation has also caught the mood, and is calling for an upsurge in funding for mathematics, in part to underpin biology.

These initiatives are bringing scientists from diverse backgrounds into biology labs. Before Odell made the switch to biology he was interested in fluid dynamics; Hood's institute has recruited George Lake, a mathematician who has worked in astrophysics and planetary science. Perhaps the biggest challenge the field faces is getting mainstream cell and molecular biologists to collaborate with these theorists and mathematicians. "The rate-limiting step will be the state of mind," predicts Albert



Hopeful signs: although the run up to the Durban AIDS conference was characterized by doubts and dissent, the meeting itself put the problems of the developing world firmly on the map.



AP

Libchaber, a theoretical physicist at Rockefeller University in New York.

Marina Chicurel

1. von Dassow, G., Meir, E., Munro, E. M. & Odell, G. M. *Nature* **406**, 188–192 (2000).

2. Alon, U., Surette, M. G., Barkai, N. & Leibler, S. *Nature* **397**, 168–171 (1999).
3. Gardner, T. S., Cantor, C. R. & Collins, J. J. *Nature* **403**, 339–342 (2000).
4. Elowitz, M. B. & Leibler, S. *Nature* **403**, 335–338 (2000).

AIDS

In and out of Africa

Three issues dominated the AIDS agenda this year: Africa, Africa and Africa. First, South African President Thabo Mbeki questioned whether HIV causes AIDS, enraging mainstream scientific opinion and, many would argue, risking the lives of millions of his citizens. Second, researchers were forced to examine a controversial theory that AIDS began in Africa with the use of contaminated polio vaccines. But, ultimately, 2000 may be remembered as the year in which the developed world finally acknowledged the enormity of Africa's AIDS crisis, and began to respond appropriately.

Mbeki's stance on AIDS was reportedly influenced by his browsing of the Internet. Facing demands to supply the drug AZT (zidovudine) to the 25% of his citizens who carry HIV — in particular, pregnant women — Mbeki discovered websites questioning the drug's safety. As he surfed, he encountered the views of Peter Duesberg, the virologist at the University of California at Berkeley who claims that HIV does not cause AIDS.

When it became clear that Mbeki was taking Duesberg and his acolytes seriously, mainstream researchers were moved to action. In July, on the eve of the XIII International AIDS Conference, held in the South African city of Durban, they published the 'Durban Declaration', reaffirming the central role of HIV in AIDS (see *Nature* **406**, 15–16; 2000). And although Mbeki has since backed

away from the debate, delegating AIDS policy to a cabinet committee, feelings among those who have devoted their careers to fighting HIV are still running high. "Many African nations look to South Africa for leadership," says David Ho of the Aaron Diamond AIDS Research Center in New York. "History will send Mbeki down in flames for his failure to properly address a disease that could ultimately kill a quarter of his countrymen."

While the response to Mbeki was consuming the energies of many AIDS researchers, others became embroiled in a controversy ignited by the British journalist Ed Hooper. In his book *The River*, Hooper argued that AIDS originated in the late 1950s, when researchers, working in what is now the Democratic Republic of the Congo, immunized local people against polio. Hooper claimed that batches of the vaccine were prepared using cultures of tissue from chimpanzees that were infected with SIV, a virus known to be the closest relative of HIV-1.

The controversy came to a head in September, at a meeting held at the Royal Society in London. The meeting heard that tests on surviving samples of the vaccine revealed no traces of HIV or SIV. And, although the case cannot be definitively closed, the sting seems to have been taken out of a debate that many researchers regard as a distraction from the pressing task of fighting the current pandemic. "Before the Royal Society meeting I

spent literally hundreds of hours on this subject, but since I returned from London I have not heard a peep from Hooper or his supporters," says Beatrice Hahn of the University of Birmingham at Alabama, who works on the molecular evolution of HIV.

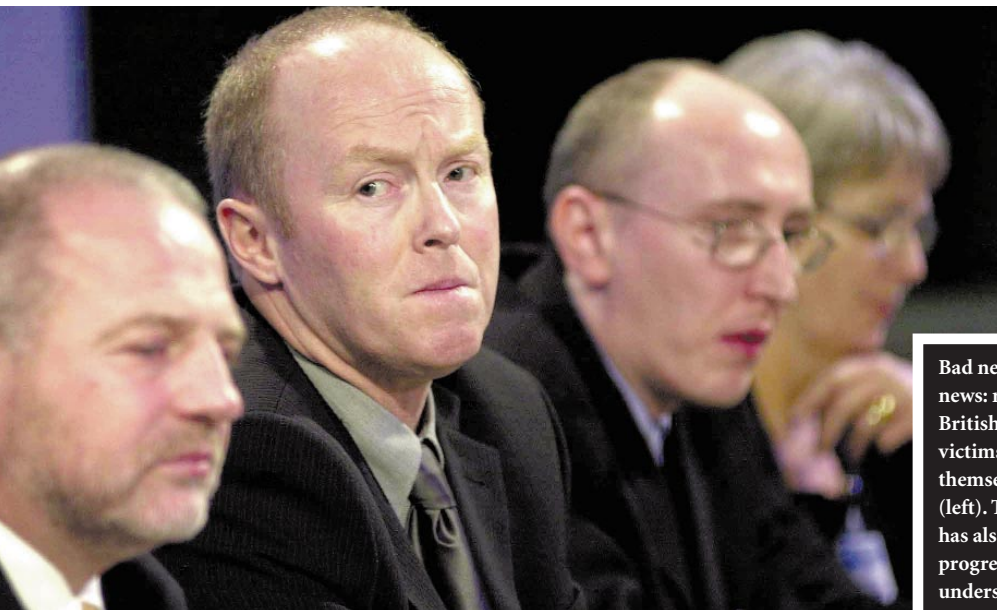
The Durban AIDS meeting, meanwhile, may come to be seen as having produced a lasting legacy. "Durban finally put the problems of facing AIDS in the developing world on the big map," says Mario Clerici, an AIDS researcher at the University of Milan. Indeed, despite the controversy he generated, Mbeki's speech at the meeting, which stressed the role of poverty in the AIDS epidemic, may have had the benefit of drawing international attention to issues such as drug pricing in developing countries.

The Durban meeting perhaps galvanized a momentum that was already under way. In January, the United Nations Security Council, for the first time in its history, had a health issue on its agenda: AIDS in Africa. In May, US President Bill Clinton released an executive order allowing developing countries to make or import generic versions of AIDS drugs without having to fear trade retaliations.

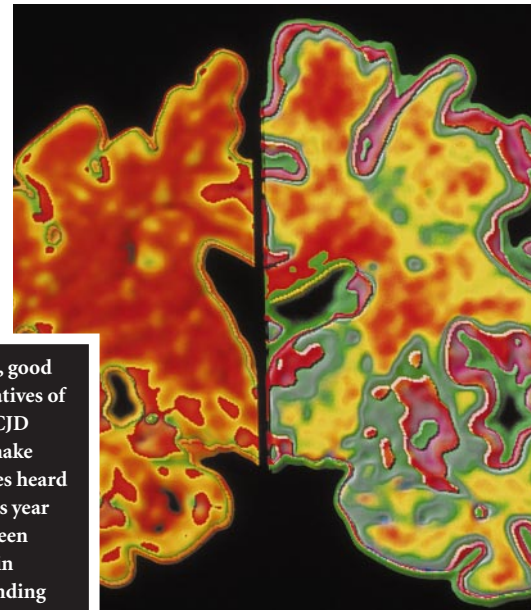
That same month, five pharmaceutical companies said they would work with the United Nations to find a strategy for HIV and AIDS in developing countries, including cheaper drugs; and Bristol-Myers Squibb launched a US\$100 million effort to fight AIDS in five southern African nations. In July, Merck and the Bill & Melinda Gates Foundation together pledged a similar amount to Botswana. And the World Bank also this year said it would put no ceiling on spending on AIDS in Africa.

The year ended with the launch of a Global AIDS Research Initiative and Strategic Plan by the US National Institutes of Health. This is supposed to give much higher

AP



STEFAN ROUSSEAU/PA



ALFRE PASIEKA/SPL

Bad news, good news: relatives of British vCJD victims make themselves heard (left). This year has also seen progress in understanding the pathology of Alzheimer's disease (right).

priority to research in and for the countries worst affected by the pandemic. Those fighting AIDS on the African front line will be watching keenly to see how the plan takes shape.

Public-health experts welcome these initiatives, but warn that much more is need-

ed if the pandemic is to be controlled. "We need a call to arms, and brave leadership," says Richard Feachem, director of the Institute for Global Health at the University of California.

Declan Butler

that prions bind to a blood protein called plasminogen² could prove significant. This

binding could form the basis of a diagnostic test, or be used to purge blood of prions. The interaction might also become a target for drug development.

Despite the shadow of vCJD, the mood elsewhere in the field of neurodegeneration — particularly in Alzheimer's disease research — is more upbeat.

The brains of people with Alzheimer's contain plaques of a peptide called amyloid- β , which many researchers believe cause dementia. Amyloid- β forms when a protein called amyloid precursor protein is broken down by the enzymes β - and γ -secretase. Over the past 12 months, studies on these enzymes have boosted the chances of developing drugs that can disrupt their activity. Papers describing the purification and cloning of β -secretase appeared just over a year ago³⁻⁵, and in October this year a team from Oklahoma produced a three-dimensional structure for the enzyme's active site⁶.

Meanwhile, researchers seeking the identity of γ -secretase were homing in on proteins found in cell membranes called presenilins⁷⁻⁹. And in September, an international team clarified the picture by producing evidence that γ -secretase is actually a complex of a presenilin and another membrane protein called nicastrin¹⁰.

While companies work to turn these advances in understanding into drugs that can disable the enzymes, one of the most surprising developments in Alzheimer's research has been the progress of a simple vaccine against amyloid- β . In 1999, the vaccine was shown to prevent plaque development in a mouse model of the disease¹¹. And two papers in this issue extend the mouse studies to show that the vaccine can partially dissolve plaques that have

Neurodegeneration

Battling the killer proteins

Scientists who work on the clinical aspects of neurodegenerative diseases know the terrible effects of these conditions all too well. Those of us fortunate enough not to have lost a loved one to Alzheimer's or Creutzfeldt-Jakob disease rarely give these killers a moment's thought.

But the release in October of the public inquiry into Britain's bovine spongiform encephalopathy (BSE) epidemic jolted some of those present out of their complacency. After government ministers and the report's authors had had their say, the stage was given to a panel from the Human BSE Foundation, which represents the families of those who have died from variant Creutzfeldt-Jakob disease (vCJD), as the human form of BSE is known.

The panel answered reporters' questions quietly, until they were asked to describe the suffering they had endured. Each had seen their relatives slowly robbed of their humanity, suffering first from depression, then from dementia and hallucinations, before their life ebbed away. One told of the final indignity of seeing his brother disposed of in a bodybag in a lime-dusted grave. At points, panel members broke down.

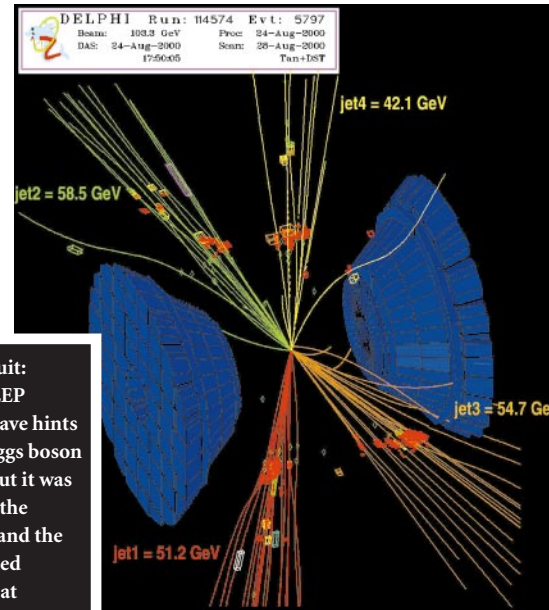
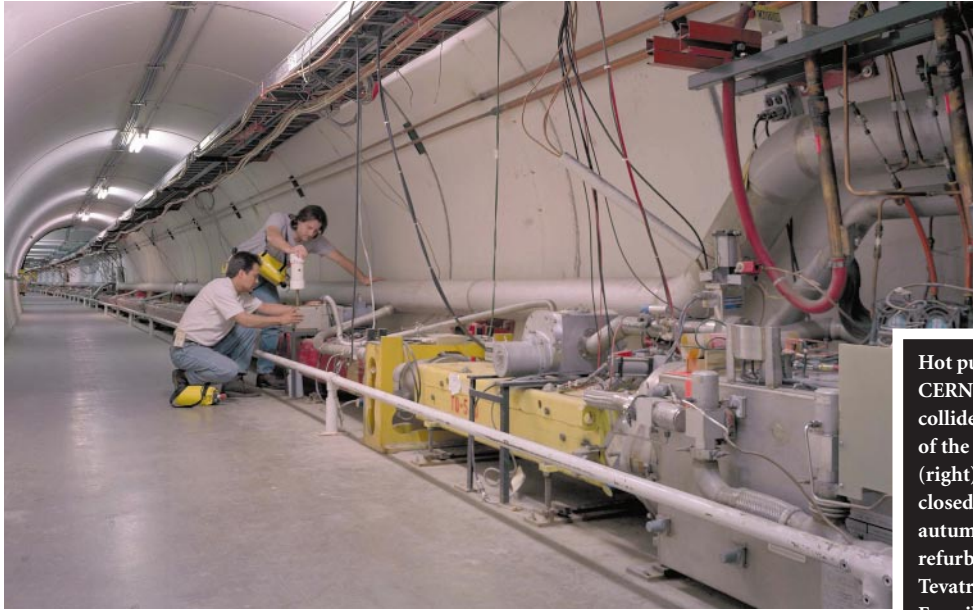
Will science one day prevent such tragedies? In the case of vCJD, the outlook is

still bleak. We know little about the infectious proteins, or prions, that are thought to cause the disease. What is more, BSE is no longer just a British problem: the disease is more widespread in French herds than was once thought, and in November Germany and Spain recorded their first cases in cows born within their borders.

The most significant unknown is the number of people who may be affected. As *Nature* went to press, there were 91 confirmed and probable cases of vCJD — the vast majority in Britain, three in France and one in Ireland. In August, researchers at the University of Oxford put the upper limit on the British epidemic at 136,000 people¹. But their epidemiological models assume that the disease will strike only in the 38% of the population with a genotype that makes them susceptible to the disease.

All the vCJD cases recorded to date fit this mould. But a similar disease called kuru, spread by cannibalism among the people of Papua New Guinea, has proved not to be restricted in this way — people with other genotypes simply took longer to succumb.

The immediate research priorities are to develop tests that can identify people incubating vCJD, and drugs to prevent the prions doing their damage. The discovery



Hot pursuit: CERN's LEP collider gave hints of the Higgs boson (right). But it was closed in the autumn, and the refurbished Tevatron at Fermilab (left), may pinpoint the particle first.

FERMILAB

already formed, and reverse some of the associated cognitive deficiencies^{12,13}.

Many researchers would have predicted that it would have been impossible to raise an immune response against one of the body's own peptides, or that the antibodies would have failed to penetrate the brain. "There were many reasons not to test out the idea," says Dale Schenk of Elan Pharmaceuticals in South San Francisco, whose team developed the vaccine. "Sometimes it pays not to think too much."

Schenk cautions that what works in mice may not work in people. The vaccine has already passed through initial human safety

studies, however, and larger trials to determine its clinical efficacy are being planned.

Peter Aldhous and Alison Abbott

1. Ghani, A. G., Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. *Nature* **406**, 583–584 (2000).
2. Fischer, M. B. *et al. Nature* **408**, 479–483 (2000).
3. Vassar, R. *et al. Science* **286**, 735–741 (1999).
4. Yan, R. *et al. Nature* **402**, 533–537 (1999).
5. Sinha, S. *et al. Nature* **402**, 537–540 (1999).
6. Hong, L. *et al. Science* **290**, 150–153 (2000).
7. Wolfe, M. S. *et al. Nature* **398**, 513–517 (1999).
8. Li, Y. M. *et al. Nature* **405**, 689–694 (2000).
9. Esler, W. P. *et al. Nature Cell Biol.* **2**, 428–434 (2000).
10. Yu, G. *et al. Nature* **407**, 48–54 (2000).
11. Schenk, D. *et al. Nature* **400**, 173–177 (1999).
12. Morgan, D. *et al. Nature* **408**, 982–985 (2000).
13. Janus, C. *et al. Nature* **408**, 979–982 (2000).

High-energy physics

When priorities collide

Change often brings regrets as well as the excitement of new beginnings. And that was definitely the case this year in high-energy physics. As the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory on Long Island, New York, detected its first particles, physicists at CERN, the European Laboratory for Particle Physics near Geneva, bade a fond, even anguished, farewell to their stalwart accelerator, the Large Electron-Positron (LEP) collider.

Nine years in the making, the RHIC reported its first collisions in June (see Back, B. B. *et al. Phys. Rev. Lett.* **85**, 3100–3104; 2000). The collider accelerates gold nuclei to nearly the speed of light in two accelerator rings. When two ions collide head on, they create a mini-fireball with a temperature of one trillion kelvin — ten thousand times hotter than the inside of the Sun.

This should recreate the conditions that

existed about ten millionths of a second after the Big Bang, before the soup of quarks and gluons had condensed to form protons and neutrons. The RHIC's main objective is to study the properties of this 'quark-gluon plasma'. Researchers at LEP reported hints of this exotic form of matter in February, but the RHIC will be able to study the plasma in detail.

For the researchers who have tended LEP, pulling the plug was always going to be an emotional wrench. But few would have predicted the drama that unfolded in September, when LEP gave tantalizing hints of the elusive Higgs boson — the fundamental particle that is thought to give other particles their mass.

For its parting shot, LEP was fired up to run at energies beyond those it had been designed for. Within weeks, physicists had recorded five events consistent with a Higgs boson with a mass of 115 giga-electron volts.

Those results bought a stay of execution until November. But to secure the claim,

more data were needed. They failed to materialize, and CERN's management decided that the evidence was not strong enough to delay the construction of the lab's next centrepiece, the Large Hadron Collider (LHC), which will occupy the same tunnel. Some physicists are still smarting over the decision.

"We felt there was enough evidence to justify running LEP again next year," says Tiziano Camporesi of LEP's DELPHI experiment.

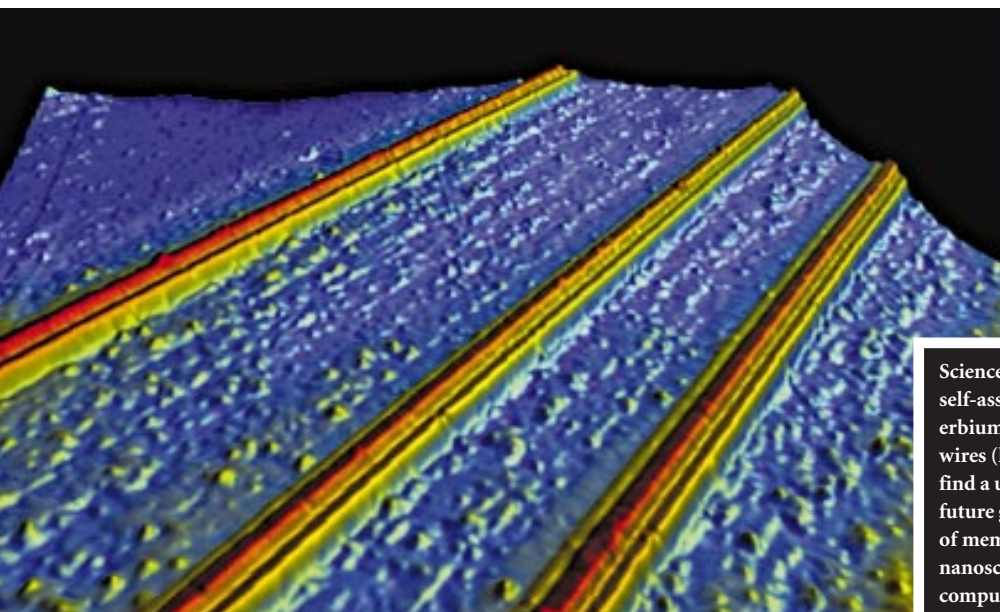
Within two years of its completion in 2005, the LHC should yield huge numbers of Higgs particles. But by then, US physicists may have stolen Europe's thunder. Fermilab near Chicago will in the spring restart its Tevatron collider, after a major refit. Although it was not designed to look for the Higgs, pinning down the signature of a particle with a mass of 115 giga-electron volts should be well within its capabilities. "Fermilab will use all its efforts and resources to solve the question," says a bitter Patrick Janot, physics coordinator at LEP. "Then CERN will look ridiculous at having missed this opportunity."

The Tevatron got another feather in its cap this year. Analysing data collected in 1997, researchers working on an experiment called DONUT reported in July that collisions between a high-energy electron beam and a tungsten target had produced five sightings of the tau neutrino, the most massive of the three types of neutrino.

This particle and the Higgs complete the bestiary predicted by high-energy physics' prevailing theory, the Standard Model, and its extension, the Higgs mechanism. But physicists will still have plenty to do when and if the Higgs is found — in particular searching for the 'supersymmetric' interactions that may unite the two classes of fundamental particles, fermions and bosons, which differ in the property of 'spin'.

Philip Ball

DELPHI, CERN



Science writ small: self-assembling erbium disilicide wires (left) may find a use in future generations of memory-rich nanoscale computers. Other groups are making microscopic machinery (right).

Nanotechnology

Molecular movers and shakers

When nanotechnology became a buzzword about a decade ago, no one was quite sure what it was. Just how the field will develop is still unclear, but the past year has seen a transformation in its ability to attract public investment. The US federal government will almost double its spending on nanotech next year, to more than \$400 million. Japan is planning a budget hike of more than 40%, and several European countries have made nanoscale research a priority. Nanotechnology looks poised to shed its science-fiction image and don the mantle of respectability.

But what opportunities should we expect to see the new funds create? The highlights of the past 12 months give some pointers. One of these is nanotechnology's potential to reinvent and revitalize chemistry. For example, chemists should have fun with nanotech's party piece, the manipulation of individual atoms using the scanning tunnelling microscope (STM).

In September, a team at the Free University of Berlin synthesized a biphenyl molecule from two benzene radicals using the STM¹. Such piece-by-piece molecule-building, although impressive, is unlikely to replace standard chemical synthesis. But the combination of nanoscale manipulation and spontaneous chemical processes has huge potential.

This was shown in July by researchers at the Steacie Institute for Molecular Sciences in Ottawa, Canada². Robert Wolkow and his colleagues used the STM to remove individual hydrogen atoms from a hydrogen-covered silicon surface. This allowed a styrene molecule to bind to the silicon, setting off a

chain reaction in which a neighbouring hydrogen was displaced, another styrene bound to the silicon, and so on — resulting in rows of molecules up to 13 nanometres long.

In a similar vein, Stanley Williams and colleagues at the Hewlett-Packard Research Laboratories in Palo Alto, California, reported in June that they had made grid-like arrays of self-assembling erbium disilicide nanowires on a silicon substrate³. They anticipate using such grids in a memory-rich architecture for a nanoscale computer. And Williams's collaborator James Heath and his team at the University of California at Los Angeles have developed another of the building blocks for such a device: molecular switches that work at room temperature⁴.

The connections in nanoscale circuits could well be made of conducting carbon nanotubes. And the discovery of a simple method for fashioning them into 'Y' shapes broadens their scope for use in electronic circuitry⁵. Conducting organic materials might open the way to a genuine molecular electronics — as was acknowledged by this year's chemistry Nobel.

The cell, meanwhile, is a ready-made toolbox of molecular machines, and biomolecular science is sure to be a big part of nanotechnology. By coupling the ability of specific biomolecules to recognize one another with manipulation using laser beams as optical tweezers, chemist George Whitesides and his colleagues at Harvard University this year explored the frontier with biology, making sculptures from red blood cells tagged with polymer microspheres⁶.

In June, a paper from Angela Belcher and colleagues at the University of Texas at Austin united protein chemistry with semiconductor technology. They created peptides that can recognize and bind to the surfaces of different semiconductors⁷. This points to the possibility of devices based on biological molecules, such as the motor proteins that power cell movement, that can assemble electronic or other inorganic structures. The exciting beginnings of such a hybrid technology were heralded in November, with the report of a nanoscale metal rotor powered by the enzyme ATP synthase⁸.

Practical applications remain years away. "Nanotechnology doesn't yet exist," says Don Eigler of IBM's Almaden Research Center in San Jose. But there is one concrete sign that nanoscale research will eventually deliver working technologies: the spin-off companies launched by some of the field's academic pioneers. In the past year, Richard Smalley of Rice University in Houston formed Carbon Nanotechnologies, which aims to commercialize the use of carbon nanotubes; and Chad Mirkin and co-workers at Northwestern University in Illinois have launched Nanosphere. This company seeks to use a system based on tiny gold particles to develop diagnostic tests that recognize particular sequences of DNA. ■

Phillip Ball

1. Hla, S.-W., Bartels, L., Meyer, G. & Rieder, K.-H. *Phys. Rev. Lett.* **85**, 2777–2780 (2000).
2. Lopinski, G. P., Wayner, D. D. M. & Wolkow, R. A. *Nature* **406**, 48–51 (2000).
3. Chen, Y., Ohlberg, D. A. A., Medeiros-Ribiero, G., Chang, Y. A. & Williams, R. S. *Appl. Phys. Lett.* **76**, 4004–4006 (2000).
4. Collier, C. P. *et al. Science* **289**, 1172–1175 (2000).
5. Satishkumar, B. C., Thomas, P. J., Govindaraj, A. & Rao, C. N. R. *Appl. Phys. Lett.* **77**, 2530–2532 (2000).
6. Holmlin, R. E. *et al. Angew. Chem. Int. Edn* **39**, 3503–3506 (2000).
7. Whaley, S. R. *et al. Nature* **405**, 665–668 (2000).
8. Soong, R. K. *et al. Science* **290**, 1555–1558 (2000).

Cloning's not a new idea: the Greeks had a word for it centuries ago

Sir—The term 'cloning' originates from the Greek word *clonos*, meaning 'twig'; *clonizo* is the verb 'to cut twigs'. A kind of cloning has been widely used in agriculture for centuries: by this process a new tree is created from an adult tree's twig without use of its seed, analogous to the cloning of mammals from adult cells.

The idea that cloning is also possible in humans also evolved in ancient times: it was realized that the principles of reproduction did not lie in the reproductive organs or seed.

Joannes Philoponus, an Alexandrian philosopher of the sixth century AD, commenting on Aristotle's writings, observed: "If someone cuts a twig from a walnut tree in Athens and plants it in Patras [200 km away], two or three years later it will bear nuts that are the same in every aspect, in size and taste and colour and every other character, with the ones from the walnut tree in Athens. ... So, if the resemblances between the plants do not originate because the seed comes from the entire body, as it was proved, but for some other reason, which [reason] he reveals later, the same applies to animals as well."

Philoponus used the word *clados* for 'twig', which is synonymous with *clonos*.

He was trying to find a reason for the resemblance between parents and children. Although he made many mistakes, he did correctly imagine the existence of some minimal part of the animal's body that contains all the information for the creation of the animal.

His thoughts were inspired by Aristotle. In his book *On Animal Generation* Aristotle proved that the information for the creation of an animal existed in all parts of the body but that, in opposition to other thinkers' beliefs, the semen did not come from the entire body in order to contain this information.

"Children are like their more remote ancestors from whom nothing has come," Aristotle wrote, "for the resemblances recur at an interval of many generations, as in the case of the woman in Elis who had intercourse with an Ethiop; her daughter was not an Ethiop but the son of that daughter was. The same thing applies also to plants."

Aristotle, many centuries before Mendel, referred to the properties of the propagation of genetic information in plants and humans. The fact that "resemblances recur at an interval of many generations" is probably Aristotle's most important observation in this regard,

implying that certain characteristics do not have to be expressed to the next generation to be perpetuated.

"If again something creates this composition later," Aristotle continued, "it would be this that would be the cause of the resemblance, not the coming of the semen from every part of the body."

Today, we know that Aristotle's "something" is DNA. Aristotle understood that something undifferentiated exists that has the potential to become a plant or an animal, and that the semen is just the carrier of that potential.

Philoponus probably used a walnut tree rather than, say, an olive tree, intentionally. The word *karyo* (nut) also meant 'testicle', then as now, and is still used today in words such as karyotype and karyokinesis.

Using Aristotle's ideas about the transmission of information from parent to child, Philoponus suggested that a kind of cloning is possible in animals. Although the use of the term 'clone' by Philoponus is a very primitive example, it is the first reported use of the word for such a process.

One would not expect a thinker from the sixth century AD to be able to clone sheep, but, as with many other modern achievements, the principle was cultivated in the minds of the ancients.

A. A. Diamandopoulos, P. C. Goudas
Chorio Romanou, 26500 Patras, Greece

Taiwan pays the price for growth, in toxic pollution

Sir—The environmental damage induced by the growth of Taiwan's technological revolution was not covered in your recent Insight feature about this country¹. A century ago, British naturalist Alfred Russel Wallace was impressed with Taiwan's natural beauty². But in the past few decades, this tiny island has evolved from agricultural backwater to global technological giant, leading to environmental disasters such as dangerously polluted rivers^{3,4}.

In 1997, the government admitted that industry had produced 146,000 tons of hazardous waste³. Using recent data from 1,000 large industrial companies, the Environmental Protection Administration estimates that Taiwan produces more than 18 million tons of technological solid waste annually; 1.47 million tons of these are considered hazardous. Only 600,000 tons are treated — the rest ends up in rivers and landfills. About 50,000 tons of toxic solvents are produced annually, 35,000 tons by firms in Hsinchu, Taiwan's high-tech industrial centre, alone¹.

In July, 100 tons of toxic solvents were

dumped into the country's second longest river, Kaoping, leaving 3 million residents (including us!) in and around Kaohsiung without drinking water for five days. In Taiwan's largest environmental criminal case to date, prosecutors alleged that the waste handler in the Kaoping affair dumped 14,000 tons of toxic solvents into river systems across the island. The government has discovered 160 illegal dump sites nationwide, of which the three most dangerous are near the river Kaoping. Furthermore, 100,000 barrels of toxic waste were recently discovered in central Taiwan, polluting the river Tatu. These examples of corporate greed are also the consequences of technological development.

Although mercury cell electrolysis was eliminated from the production line in 1989, Taiwan has accumulated mercury waste of about 100,000 tons, as estimated by the Industrial Development Bureau; an illegal mercury-tainted dump has recently been found in Hsinfeng town near Hsinchu. Taiwan currently has just one secure landfill, in Kaohsiung. It cannot handle all the toxic waste it produces, hence the government is seeking cash-starved countries that will dispose of it for a hefty fee. In a high-profile case last year, Cambodia returned 2,700 tons of Taiwan's mercury-laced waste after several deaths near the disposal site.

During the financial year 1999, Taiwan spent a large sum on national defence (20.5% of the national budget), science and education (15.8%), and economic development (13.5%), but environmental protection remained a low priority (1.5%)⁵.

Taiwan's technological miracle has taken place at the cost to our society of excessive toxic waste and pollution. The country now needs a profound reorientation in its attitude to the environment, with the help of the country's small but growing green movement. AD
Govindasamy Agoramoorthy*,
Minna J. Hsu†

*Department of Wildlife Conservation, National Pingtung University of Science & Technology, Pingtung, Taiwan

†Department of Biological Sciences, National Sun Yat-sen University, Kaohsiung, Taiwan

1. Swinbanks, D. & Cyranoski, D. *Nature* **407**, 415–426 (2000).

2. Wallace, A. R. *Island Life* (Macmillan, London, 1880).

3. 1999 Republic of China Year Book (Government Information Office, Taipei, 1999).

4. Hsu, M. J. & Agoramoorthy, G. *Environ. Conserv.* **25**, 288–289 (1998).

5. Republic of China budget for 1999 in Chinese, www.dgbase.gov.tw/account; the 1998 budget in English at http://www.dgbase.gov.tw/english/dgbas_e3.htm.

Erratum The headline "Island-hopping invaders hitch a ride with tourists in South Georgia" (*Nature* **408**, 637; 2000) should have read "in the Southern Ocean" instead of "in South Georgia". *Nature* apologizes for this error.

A diaspora and its blessings

Jewish academics exiled from 1930s Germany rejuvenated science elsewhere.



Hitler's Gift: Scientists Who Fled Nazi Germany

by Jean Medawar & David Pyke
Metro Publishing: 2000. 268 pp. £20

Walter Gratzer

"My country, right or wrong" was, in G. K. Chesterton's view, a maxim unworthy of a true patriot — no better, he thought, than "my mother, drunk or sober". Yet it encapsulates the stance of most German academics after Hitler came to power in 1933. Some ardently welcomed the radical new regime, and some merely imagined that it could be educated out of its more uncouth habits; others were persuaded that it would not last anyway. But many more perceived only the opportunities for advancement that would follow from the ejection of Jews and political undesirables.

There is only one recorded instance of a principled refusal to slide into the shoes of a dispossessed colleague. Jean Medawar and David Pyke relate the story of Otto Kroyer in their vivid account of the scientific diaspora out of Germany in the years before the Second World War. But they do not mention another remarkable case, that of Curt Kosswig. A professor at the Technical University of Braunschweig, Kosswig was an early adherent of the Nazi Party (NSDAP) and a member of the SS. He was, moreover, a geneticist, specializing in the genetics of race — at the time a highly favoured discipline. But he refused to endorse the infamous racist Nuremberg Laws and spoke out against the persecution of the Jews. He was expelled from the SS and evaded arrest only by a hasty exit to Turkey. There was little other activity in Germany at the time to engender faith in human nature.

The exodus of German, Austrian and, later, Italian academics has been likened to the flight

Climate of fear: the prejudice that burned books drove many scientists from Germany.

of scholars from Constantinople in the fifteenth century. German science, and scholarship generally, had led the world in the preceding years. Of the 100 Nobel prizes for science awarded between 1901 — their first year — and 1932, no fewer than 33 went to Germany, eight of them to (largely assimilated) Jews.

The Nazi regime was unperturbed by the injury the Jews' departure was visiting on science. "If science can't do without Jews," Hitler instructed Max Planck, the president of the Prussian Academy of Sciences, "then we shall have to do without science for a few years." But, in truth, the decline of German science resulted not so much from the disappearance of so many of its leaders, for plenty of first-rate practitioners still remained, as from the degradation of learning by ideological imperatives, and the ravages of the *Führerprinzip* which propelled party non-entities into the most influential university positions. This in turn led to a calamitous decline in the recruitment of able students, with consequences that endured for decades.

For Britain and America there was a rich windfall. "Hitler is my best friend," declared the director of the Institute of Fine Arts in New York. "He shakes the tree and I collect the apples." The émigrés rejuvenated science in Britain and the United States and, above all, launched the United States on the path to scientific primacy. English alone, and not German, is now the universal language of science.

The most heartening theme running through Medawar and Pyke's story is the manner in which a group of academic luminaries in Britain reacted to the events that started to unfold in Germany in 1933. The heroes are most notably the economist William Beveridge (later director of the London School of

Economics and architect of the welfare state under the postwar Labour government), the physiologist, Nobel laureate and Member of Parliament for Cambridge University A. V. Hill, and the biochemist and Nobel prize-winner Frederick Gowland Hopkins. Together, they founded the Academic Assistance Council in 1933. They struggled to persuade a recalcitrant Home Office to allow German academics into the country, and then scraped together salaries for the new arrivals.

They saw far more clearly than the politicians what the exodus portended. Hill denounced the activities of the egregious Johannes Stark, champion of the *Deutsche Physik*, which sought to expunge from science such alien pollutants as relativity and quantum theory, in a fierce exchange in *Nature* — and indeed, the mathematician G. H. Hardy did the same, directing his magisterial scorn at the mathematical Nazi ideologue Ludwig Bieberbach. Bernard Katz, Hill's protégé in the physiology department of University College in London, described his patron as "the most naturally upright man I have ever known".

David Pyke has added a personal reminiscence. Dismayed at the plight of the refugee doctors while still a medical student in Cambridge and editor of the student magazine, he addressed an appeal to Hill; notwithstanding the other and more important matters he must have had on his mind, Hill wrote a long and courteous reply, which Pyke reproduces in his preface to the book.

The savants, of course, were fortunate in finding such champions and protectors. But spare a thought here for the desperate majority of fleeing Jews, who, enjoying no such patronage, came up instead against the unyielding British Home Office policy of "no admittance without financial guarantees".

In 1940, almost all German and Austrian Jews in Britain were classed as enemy aliens, as were many Italians — mostly chefs and waiters, long resident in the country. After much debate, Winston Churchill issued instructions to "collar the lot". The story of the internment camps on the Isle of Man and the deportation of so many of the captives to Canada and Australia has been often told and is well surveyed here. Max Perutz — who made the journey to the Isle of Man, thence to Canada and finally back to England to embark on an ill-starred war project with David Pyke's father, the ingenious and eccentric inventor Geoffrey — has written that the commandant of the internment camp mused aloud on the curious circumstance that so many of these frightful Nazis were Jews.

In their chapter, "Those who stayed",

book reviews

CHARLES & JOSETTE LENARS/CORBIS

Medawar and Pyke reflect on the scientists who chose to remain in Germany and in many cases help to advance the German war effort. Here, for the most part, is human nature at its least attractive. Werner Heisenberg, who had been pilloried as a "white Jew" — an adherent of the alien pseudoscience of relativity and quantum theory that was polluting the purity of Aryan thought — was crass enough to tell the Jewish émigré physicist Francis Simon after the war that "the Nazis should have been left in power longer, then they would have become quite decent".

When the friend and collaborator of Max Born, Pascual Jordan, who had transformed himself into a venomous Nazi and anti-Semite, appealed to Born for a testimonial to present to the denazification tribunal, Born merely sent a list of relatives who had perished in Auschwitz.

The most admirable of the physicists who stayed was the courageous and incorruptible Max von Laue, the only German to retain Einstein's esteem. There were others: Fritz Strassmann, who sheltered a Jewish fugitive through much of the war at risk to his life, is one who might have earned a mention.

Much of *Hitler's Gift*, however, is taken up with potted biographies of protagonists, among them several of the authors' friends. These reminiscences, affectionate and often moving, are the most compelling part of the book. They include many, but not, as Medawar and Pyke suggest, most, of the leading figures. Even some Nobel laureates (the authors' touchstone) are absent — Konrad Bloch and Otto Stern are two — and representing the Italians is Enrico Fermi, but not Rita Levi-Montalcini or Salvador Luria. Of the mathematicians, Richard Courant is here, but not Hermann Weyl or Emmy Noether.

There is a fairly generous sprinkling of errors. For example, Otto Warburg received his Nobel prize for work, not on muscle, but on respiratory enzymes; the first X-ray diffraction experiment was performed by Friedrich and Knipping, not Kripping; it was Arnold (not Arthur) Sommerfeld, and so on.

But it would be ungrateful to cavil. Jean Medawar and David Pyke have told their life-affirming tale very persuasively. They show with what devotion and loyalty most of the émigrés repaid their debt to their adoptive countries. Not a few felt that their lives had been enhanced by their enforced transplantation into a new and more open culture. It was Ernest Rutherford who said that scientists never grow old, for the innocent pleasures of discovery will last them a lifetime. For so many of the European exiles, hounded out of the country they had loved, served and often fought for, science remained their hope and consolation. When the physiologist Otto Loewi — who shared the Nobel prize with Henry Dale for the discovery of chemical neurotransmission — was seized by the Austrian Storm Troopers, he persuaded a turnkey

to post to the journal *Naturwissenschaften* a card bearing a scrawled summary of his most recent laboratory results. Loewi then felt, according to Medawar and Pyke, an "in-describable relief", transcending his concern for his incarcerated wife and children, that his discoveries would not be lost if, as he fully expected, he was murdered. (He was saved by the intervention of Henry Dale.)

To conclude, *Hitler's Gift* is worth the money for a characteristically limpid foreword by Max Perutz, one of the most celebrated of all the émigrés, and for an appendix comprising part of Perutz's splendid memoir, taken from *The New Yorker*, of his wartime experiences in exile. ■

Walter Gratzer is at the Randall Centre, New Hunt's House, Guy's Campus, King's College, St Thomas Street, London SE1 1UL, UK.

Tuck in, and enjoy!

The Cambridge World History of Food

edited by Kenneth F. Kiple & Kriemhild Coneè Ornelas
Cambridge University Press: 2000. 2 vols.
2,153 pp. £95, \$150

Anthony Blake

By any standards, this is a big book. Having agreed to review it within a relatively short time, I was not immediately sure where to start when all 5.4 kilograms of it arrived on my desk. Nevertheless, once I had challenged its 2,153 pages by looking for what I judged to be suitably obscure topics and found most of them, I was enthused to get on with working my way through in a more methodical way.

The two volumes, appropriately decorated with the art of Giuseppe Archimboldo, have required ten years to compile contributions from almost 160 specialist authors. The work is the result of the Cambridge History and Culture of Food and Nutrition Project, launched in 1990 when the earlier *Cambridge World History of Human Disease* project neared completion. From this origin, it is not surprising that these volumes bring together the intertwining relationships between the development of human societies, the foods they have become dependent on and how these relate to human health and well-being.

The editorial board has done an excellent job in putting the articles together to make a coherent and well-structured whole. As the editors explain in their preface, these volumes were never intended to be an encyclopaedia, but rather a collection of original essays. The sheer diversity and scope of the articles has inevitably led to some overlap of subject matter, but in my view this adds to the books' appeal, bringing a



Stuff of life: tortillas being prepared by Huastec women, from a fresco painting by Diego Rivera.

range of emphasis and perspective.

There are two implicit threads running throughout the work. The first is the globalization of food in the past 10,000 years, since we ceased to be hunter-gatherers and became herdsmen and farmers. The second thread follows the opportunities and challenges that technological change has presented during the repeating cycles of improved food supply, resultant population growth and further demands for food.

Volume 1 is devoted primarily to food itself and the nutritional and health issues related to diet. It starts with a review of what we know from archaeological evidence about the foods of hunter-gatherers and the state of their health. There is evidence that small groups of humans living off the land could and did enjoy well-balanced and nutritious diets, and may well have been healthier than their more civilized descendants.

Sections on staple foods and dietary liquids together make up more than half of this volume. They cover not only the obvious staples such as cereal grains, livestock and vegetables, but many less well-known foods, such as algae and alpacas. This section was fascinating and I found many facts that I will no doubt raise over dinner — for example, 'Did you know that by collecting sun-dried grasshoppers from the beaches of the Great Salt Lake in Utah you can obtain more kilocalories of food energy per hour than from any other known collectable resource?'

The remainder of the volume is devoted to health issues, such as the diseases that result from deficiencies in vitamins and essential nutrients in the diets of people in less developed countries.

Volume 2 concentrates on the sociological aspects of food and its production. It starts with historical issues and the transition from hunter-gatherer societies to agricultural communities, particularly in the Fertile Crescent of the Near East. A large number of articles follow which discuss the development of food resources and the human diet by geographical region, with an unparalleled diversity of information on the history of food in different cultures and countries in one volume.

It is not surprising that the history of nutrition and food-related health is well covered, and this includes interesting articles on the psychology of food choice, food fads, prejudices and taboos. Initially, I thought that food within religion had been overlooked, but this is not so.

A significant part of this volume concerns the preoccupations of modern society with food and health issues. It covers the activities of governments, international organizations and lobbyists in their attempts to influence our diet, and, perhaps unsurprisingly, there is a bias towards the preoccupations in the United States, since this must be the major target market for the book. Similarly, the discussion on policies relating to dietary advice is primarily focused on the American consumer.

Issues of food safety and the impact of biotechnology are included in a five-page article, but no doubt the editors had to make a decision about when to finalize the deadline for submissions, because the current controversy surrounding the genetic modification of foodstuffs missed any mention. The final article in this section neatly brings the book full circle, debating the question of palaeolithic nutrition and modern health, and whether our biological evolution has been able to cope with the enormous changes in our diet in the intervening 40,000 years.

The 173-page "Historical Dictionary of the World's Plant Foods" is probably the section of the book that will be thumbed most. It is perhaps a pity that it does not also cover foods of animal origin, but again the editors had to draw the line somewhere. Indeed, a book that sets out to cover such an enormous subject must inevitably have omitted some specialist issues. For example, the subject of cooking and the development of cooking equipment is hardly covered, and Benjamin Thompson, who invented the cooking range, receives only a passing acknowledgement.

From my own interest in food flavour, I was a little disappointed that the topic of 'flavourings' was limited to six pages on herbs and spices, but this is hardly a real point of criticism.

Overall, this is an excellent book both for reference and for general interest. It is well written, readable, and most of the authors

include extensive bibliographies. The indexes of "Latin Names", "People's Names" and "General Subjects" make information on specific topics easy to find.

Although the sheer weight of the two volumes means this is probably not a book for bedtime reading, it will nevertheless make a perfect Christmas present for anyone interested in food — and will be guaranteed to keep them occupied for many days, if not weeks.

Anthony Blake is at Firmenich SA, 1 route des Jeunes, CH 1211 Geneva 8, Switzerland.

Blood, gore and holey stomachs

Sick! Bloody Moments in the History of Medicine

by Gael Jennings

Franklin Watts: 2000. 69 pp. £8.99

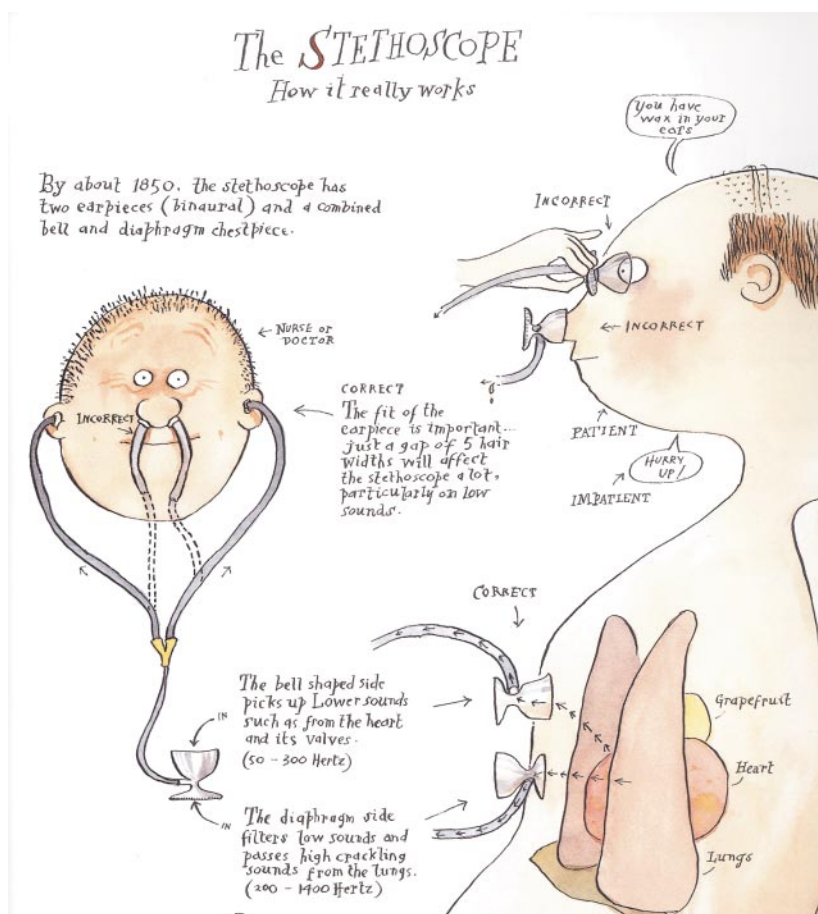
W. F. Bynum

As W. H. Auden once observed, everyone loves the smell of his (or her) own farts. Children love the sound as well, so what better way to teach a subject such as the history of medicine than by adopting a scatological perspective?

As its subtitle suggests, *Sick!* offers an episodic history of medicine — blood, guts and all. Well, not quite all. There are gangrenous legs, guts hanging out, all manner of horribly illustrated skin diseases, and even a diagram of a rectum, but no genitals in sight, despite the book's preoccupation with AIDS and other sexually transmitted diseases. The illustrations, by Roland Harvey, are striking, much like a scary Richard Scarry, but without his subtlety and, ultimately, interest.

Obviously designed for today's readers, who are encouraged to buy the book of the television programme, Gael Jennings' book pretends to be a CD-ROM. The young, bespectacled heroine Mabel receives an alternative to B-movies through her letter-box in the form of a CD-ROM labelled "The Guts of Human Life". Desperate for anything to save her from daytime TV with her geriatric dog Max, Mabel switches on and watches. What she finds is a history of medicine from Ambroise Paré to the genome.

Paré is the first modern doctor, we assume, not because, as he claimed, he treated and God healed, but because he came from humble stock, wrote in the vernacular, used ointment rather than boiling oil on battle wounds, and fathered a late, large family. This chronicle describes anaesthesia after Paré, then goes back to the anatomical



Take *this* medicine with a pinch of salt: one of the book's many illustrations by Roland Harvey.

and physiological work of Andreas Vesalius, William Harvey and their kind. Other vignettes describe Renè Laënnec's invention of the stethoscope; William Beaumont's experiments on Alexis St Martin, whose injury to his stomach left a hole large enough for Beaumont to monitor digestion as it was occurring; the development of blood transfusion; and Christiaan Barnard's first heart transplant in a human.

Jennings reserves her longest section for infectious diseases, in which we are treated to the doings of Anton van Leeuwenhoek ('Tony' to his friends, who include Jennings' readers), Ignaz Semmelweis ('Iggy') and Joseph Lister ('Joe'). Florence Nightingale, Edward Jenner, John Snow and Louis Pasteur also figure, but with less familiarity. William Halsted's introduction of rubber gloves in the operating theatre, partly out of concern for the rough, reddened hands of his nurse (and fiancée), offers the opportunity for a digression on the entry of women into medicine. The episode is rather spoilt by the misspelling of Halsted's name and place of work (here dubbed John Hopkins).

This being a modern book, it ends appropriately ambiguously, with the uncertain prospects of gene transplants and human clones. And Mabel's computer having crashed (or at least violently ejected the CD-ROM), she wisely goes out to play with Max, contemplating whether medicine is a suitable profession for her.

Anything that encourages children to think about science and medicine, and their history, is to be welcomed, and *Sick!* is a brave first book. Even though it's a work about heroes and heroines, it is not as relentlessly presentist as comparable books would have been 20 years ago. There are some problems, however. The computer conceit is clumsy and doesn't work very well. The book is too elementary for my son (aged 30), and I think too advanced for my grandsons (aged 1 and 4). It seems to be aimed at the 8–10 age group, although with a lot of technical words thrown in. The level of historical accuracy is just about acceptable, but there are too many careless mistakes, and I cringed when the origin of the "weird practice" of blood-letting was traced to our hairy prehistoric ancestors watching hippopotamuses bleed themselves.

This book set me thinking about the tension that exists between drama, good historical stories and accuracy, and it reminded me that in history (as in science), questions can crop up at any time throughout our lives, but the answers need to become better. The questions posed by *Sick!* are better than the answers it offers. ■

W. F. Bynum is at the Wellcome Trust Centre for the History of Medicine, University College London, 24 Eversholt Street, London NW1 1AD, UK.

Science in culture



Vaulting space

The construction of vaulted roofs through the ages.

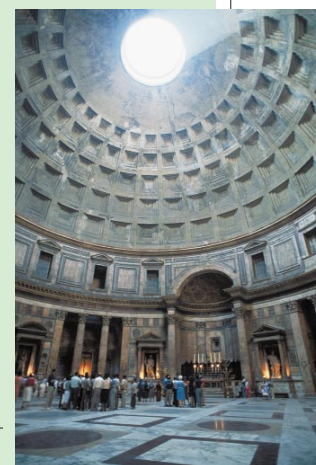
Beams make roofing easy, but clear spans are limited by the capacity of materials to take a load in tension. Stone, excellent in compression, is not very good at that: the Parthenon (447 BC) may be Greek civilization's most perfectly proportioned building, but its architraves of Pentelikon marble span just 2.5 metres.

The challenge of creating spans increases by imposing a curve: vaulted space looks lighter and more elegant — but it is trickier to construct. Only the invention of concrete made vaulting easier. The pinnacle of this Roman achievement still stands in the heart of Rome's old city. The Pantheon's bold dome, spanning 43.2 metres, consists of five rows of square coffers of diminishing size, converging on the stunning, unglazed central circular opening, the oculus. The dome's most intriguing property is its vertically decreasing specific mass, achieved by using progressively thinner layers of concrete made with lighter aggregates.

No pre-industrial builders ever topped this span — but one of the signature buildings of the Renaissance came close to the Pantheon's vault in size, and rivals it in its ingenious construction. In 1420, Filippo Brunelleschi — a prominent Florentine painter, sculptor and architect — won a competition to build an eight-sided vault without exterior buttresses over the Santa Maria del Fiore, the city's cathedral (duomo). He proposed to do so without using any wooden armature, by laying bricks in herringbone courses between a framework of stone beams, and by building "one vault inside and the other outside so that a man can walk upright between them". As the artist and Renaissance historian Giorgio Vasari wrote in 1550, the trustees of the duomo and the workmen thought "that this method was that of a madman". But the cupola, with an inner span of 41.5 metres, was completed in 1436.

Only the new materials of the nineteenth century made a rapid, and elegant, difference: the combination of cast-iron truss work and inexpensive plate glass enabled sizeable vaults to be built over greenhouses sheltering tropical trees in light-filled interiors of temperate Europe — and much bolder iron–glass vaults

Britain's Millennium Dome (above) and the Pantheon in Rome.



to cover the smoky railway stations of large cities.

Steel became the twentieth century's principal structural metal, and it is now combined with aluminium, plastics and composite materials to vault spaces of unprecedented size — but for a different worship. Modern secularism has been erecting its vaulted monuments almost exclusively for entertainment. Houston's Astrodome, completed in April 1965, was the world's first all-weather domed stadium. Its steel skeleton — a lamella frame incorporating trussed beams arching towards the centre and braced in a diamond pattern — supports a roof made of cast acrylic skylights with a clear span of nearly 193 metres. Atlanta's oval-shaped Georgia Dome has a Teflon-coated fibreglass roof whose long axis spans 225 metres. Many cities now have similar secular cathedrals for sports, entertainment, exhibitions and shopping.

Recent vaulting records go to a new airport and to a dome whose *raison d'être* appears to be quasi-religious. Hong Kong's Chek Lap Kok airport terminal, now the world's single largest building, covering more than 500,000 square metres, is vaulted in a spectacular, light-suffused style by a modular waveform roof. Its segments are constructed from I-beams forming parallelograms joined together in 36 × 36-metre vaults supported by slender beams, and daylight reflectors diffuse the subtropical sun on millions of travellers.

London's Millennium Dome is not particularly tall (just 50 metres), but its shallow curve — fashioned from coated fibreglass and suspended from twelve 100-m-high steel masts held by a net of high-strength steel cables — is 320 metres in diameter. This is an order of magnitude larger than the largest Renaissance vaults — but whether the dome above the Thames will rival the longevity of Brunelleschi's creation is another matter. ■

Vaclav Smil is at the University of Manitoba, Winnipeg R3T 2N2, Canada.

Temptations of the tree

A perennial image of life, history and enlightenment.

Geir Hestmark

Children seem to know that we belong in the trees — that climbing is the natural thing to do and that, safe up there among our primate relatives, there is a suitable branch for everyone.

At this time of the year, many people the world over bring a Christmas tree into their living room to celebrate life. The tree is one of the most powerful images in human thought and worship, a feature of human environments from taiga to rainforest, and a symbol of persistence, fertility, life, descent, destiny, purification and strength, a vertical link between the Earth and the heavens, a place to seek knowledge.

Siddharta Gautama experienced Enlightenment under the banyan tree (*Ficus benghalensis*), and became the Buddha. The seven-branched candlestick of Judaism, the menorah, symbolizes the expansion and illumination of consciousness in the image of a tree. According to the *Hávamál*, Odin, the ruler of Norse and Germanic gods, hanged and sacrificed himself for nine nights in “the tree, of which no one knows the origin of its roots” to learn the secrets of runes, language, wisdom and action. St Justin Martyr said that the Lord “reigned from the tree”, meaning both the tree of the Cross and the Tree of Life. In Eden, man was denied the fruits from the Tree of (eternal) Life to prevent him from becoming entirely like God after eating the fruits of the Tree of the Knowledge of Good and Evil.

The idea of history as a tree dates back at least to the twelfth-century mystic Joachim of Floris, who conceived the Tree of Life described by St John in the Book of Revelation as a continuous unfolding of three ages: of the Father, the Son and the Holy Spirit, under the reigns of Law, Church (Faith) and finally Love. Uprisings against the Church followed when common people asked for the speedy realization of phase three. The three-phase template of history was inherited by the German romantics and their materialist and Marxist descendants, as well as by the positivist philosophy of Auguste Comte.

From the late Renaissance, noble families in Europe had their pedigrees drawn in the form of trees — genealogical trees. By a giant leap of scientific imagination and analogy, these family trees became the prototype for a new theory of humanity's place in the Universe: “I believe this simile largely speaks the truth,” wrote Charles Darwin in *The Origin of Species* (1859), imagining evolution as a tree branching through time.

In 1776, the German–Russian naturalist Peter Simon Pallas proposed in his *Elenchus Zoophytorum* a systematic arrangement of all organisms in the image of a tree; and in 1861 Heinrich Bronn presented such a tree based on fossil evidence. But it was Ernst Haeckel, in his monumental *Generelle Morphologie* (1866), who drew the first tree of the common descent of all life on Earth. Haeckel also coined the term ‘phylogeny’ from the Greek words for tribe or race and origin, and presented phylogenetic trees for all major groups of organisms.

One might expect a German to draw thick oaks and beeches, a veritable Teutoburger Wald where the legions of Rome would lose their nerve among the mighty trunks. But Haeckel's trees look more like slender kelps — after all, he was mainly a marine biologist — that allowed him to show more detail. The German entomologist Willi Hennig subsequently created cladistics (from the Greek word for branch), a method

of taxonomy emphasizing the branch-splitting sections of evolutionary trees.

Phylogenetic trees are common in today's scientific journals, but there it is seldom realized how speculative they are because they look so real. This rhetorical power was significant in the popularization and triumph of evolutionary theory. Yet phylogenies are only sketches of historical hypotheses, constructed from imperfect historical evidence: fossils; morphological and anatomical similarity; biogeographic patterns; and, recently, comparison of different molecular sequences. Some phylogenetic trees are certainly more probable than others, but the inherently imperfect nature of the evidence seems to guarantee that we will never be able to reconstruct, except perhaps by accident, the true phylogeny of life on Earth.

Yet the need for belonging and fitting in, for creating order and enlightenment, ensures that we will spend new millennia searching downwards for our roots and upwards towards our branch in life. What better ending then, than to quote the final lines of Patrick White's epic novel *The Tree of Man* (1956): “So that in the end there were the trees. The boy walking through them with his head drooping as he increased in stature. Putting out shoots of green thought. So that, in the end, there was no end.” ■

Geir Hestmark is in the Department of Biology, University of Oslo, 0316 Oslo Norway.

We will spend new millennia searching for our roots and upwards towards our branch in life.



Phylogenies evolved from genealogies, such as this one from seventeenth-century Germany.

Monolith

With apologies to Arthur C. Clarke.

Ian Stewart and Jack Cohen

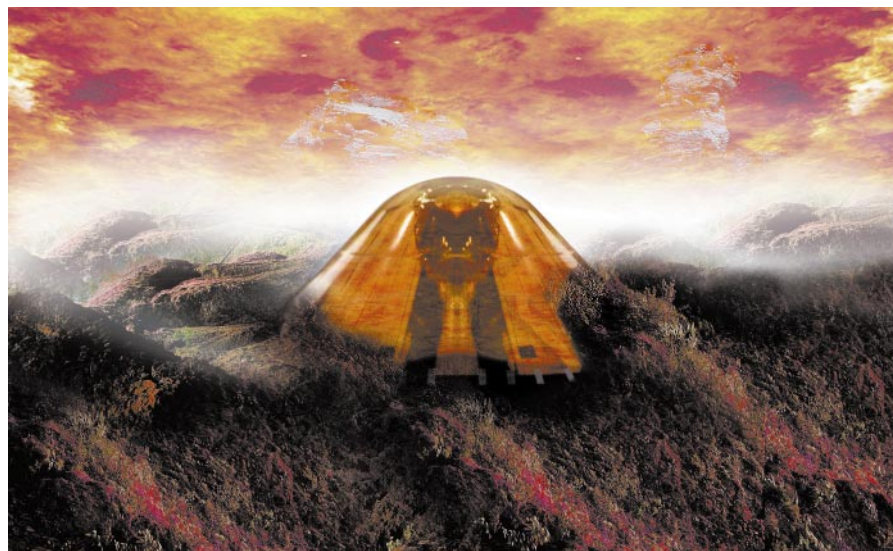
At my age, you get nervous. Most of your friends have already discovered their role in life — male, female, caretaker. If you are a late developer, though, you can't help worrying. So much depends on chance — you lie in your bed of oxy-hydride slush, wondering which role fate will select for you. Will you risk the danger of the skies to chance on an unencumbered female, arresting your development at malehood? Or will you wander your entire adolescence, alone and unloved, until you grow into a fat, gravid female? Or will you squander your heredity in the comfort of the slush-marshes, until one day you discover a clutch of eggs, settle down, and become a caretaker?

One part of me once wanted to fly free, to risk the terrors of the open skies — even to make the perilous ascent to the ceiling of warm rock that roofs the world, keeping it safe from the fires above. In past ages, belief in the overhead fires was little more than superstition, but modern science has shown that the fires really do exist. We have even observed molten rock dripping from cracks in the cosmic ceiling, to cool into grotesque stalactites as it encounters the fluid of the upper sky. Ours is an informed cosmology — the immense, finite-but-unbounded biolayer, sandwiched between the eternal fires of the rock ceiling and the indescribable cold of the undermarsh icefields. There are even savants who claim to know the total volume of the universe.

To fly, to become sexed — once, I thought to make the attempt. With tremulous strokes of my flukes I abandoned the slush-tower forest for the darkness of the liquid skies. I say 'dark', but that is a conventional exaggeration, for everywhere the skies are lit by the glow of living creatures in myriad colours — an ever-changing panorama of subtle communication in languages as yet not understood.

And that was my undoing. My confidence, I thought, was secure: somewhere in the distant heights was the rocky roof of the universe, and I would accept no lesser goal. Until I saw the gigantic, glowing mass of an engulfer hesitate, sniff the currents, and then — I still dissolve at the very recollection — turn towards me, and commence its hunting gyrations.

I was terrified. I fled. It followed. In a moment of unparalleled horror, I felt it engulf me — and pass by, suddenly revealed



as a sham. I had fled from nothing more fearsome than a shoal of nimmows, mimicking the bioluminescent spectrum of the most feared predator of our skies. Chastened, I recognized that the path of parenthood was for braver souls than I. Settling back into the homely contortions of slush-towers and polyolith-patches at the Bottom of the World, I knew myself doomed to become a caretaker. After that, my thoughts revolved solely around the size and proportions of the egg-clutch upon which I would eventually imprint. My greatest ambition was to assist in the hatching of the most numerous and most perfectly formed squablets that the universe has ever seen.

Although I continued to delude myself into the belief that I might yet regain the courage to become sexed, subconsciously I knew that my time was running out. Many times I passed the glowing lure of an egg-mass, until, one day — inevitably — I blundered upon one that smelt of my family. At once I was entranced, ensnared — lost. I tended my clutch with single-minded devotion. I flew endlessly around my eggs, fanning them with my dorsal fins to enhance the flow of nutrients. I chased off sundry predators, my fears now utterly subjugated to the hormonal imperatives of my caretaker role.

I was happy.

Then, tragedy struck. I still cannot understand it. I do not believe it can be of this universe, yet this universe, by definition, constitutes all that there is. The Monolith is a transcendent mystery, which would be wonderful had it involved anybody's brood but mine.

This is what happened. Over many cycles

I had assembled a most marvellous bed of decaying organics, ready for the implantation phase of my squablets' nurture. I painstakingly scraped away the patches of polyoliths that disfigured the serenity of the locale. Though I say it myself, it was the most perfect carrion-midden ever prepared by a doting caretaker. And then — barely a cycle before implantation was due — disaster struck.

I was rearranging fronds of putrescent geloids to bring my masterwork even closer to the pinnacle of protectiveness when — without warning, with a reality that was total and brutal — something indescribably awful caused the entire slush-marsh to shudder. Then some ghastly thing poked up from the marsh, right through the middle of my carrion-midden, ruining my life's work. At first, I thought it some rapidly-growing species of polyolith. But as the sediment settled and the sheer weirdness of the thing began to impress itself on my vision, I saw that it was all of one piece — not poly, but mono.

The Monolith is still there. Occasionally, parts of it move. What it is, I do not know. Perhaps its most inexplicable feature is its markings, which scholars braver than I have since delineated by the light of their own protuberances. I record them here for your enlightenment, in the hope that someone cleverer than I might decipher their meaning:

EXEDITION
EUROPA
VSAN

Ian Stewart and Jack Cohen's science fiction novel *Wheeler* is published by Warner Aspect. It will appear in the UK in the auspicious year 2001.

Model behaviour

Paul F. Chapman

Trials in mice of a possible vaccine for Alzheimer's disease show that it reduces the behavioural defects and the brain damage seen in the disease. As promising as these results are, a human vaccine remains a long way off.

The prospect of testing treatments for Alzheimer's disease is daunting. There are huge uncertainties about who is at risk and how the disease will progress in individual cases, so treatments might spend years in clinical trials in humans before their effects can be assessed with confidence. Good animal models of Alzheimer's disease are therefore essential as a first testing ground for potential treatments, and several mouse models have recently become available (Box 1; reviewed in refs 1, 2). Last year, one of these was used to test the effects of immunization with the β -amyloid peptide³, deposits of which in the brain are a hallmark both of patients with Alzheimer's disease and of these mouse models. The upshot of the vaccination was that the formation of β -amyloid deposits in the mice was significantly reduced. But it remains unknown whether the vaccine could also stave off the cognitive losses that make Alzheimer's disease so devastating.

Two papers in this issue^{4,5} (pages 979 and 982) raise hope that it can, at least in mice. And, together with a third paper⁶ (page 975), they suggest that appropriate animal models can be used to test the effects of potential treatments on both the brain damage and the cognitive losses caused by Alzheimer's disease. Chen *et al.*⁶ provide evidence of learning defects that depend on age (and are associated with increasing accumulation of β -amyloid peptide) in one of the main mouse models of Alzheimer's disease⁷. The authors show that these defects can be distinguished from age-independent deficits, but only by careful experimental design and analysis.

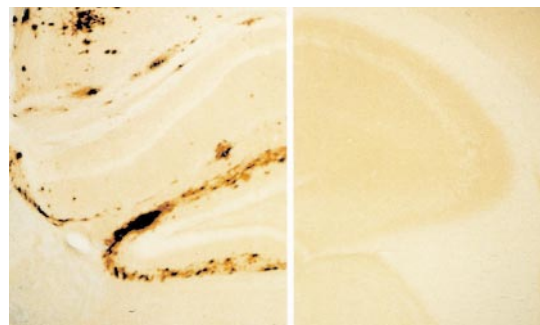
Meanwhile, Janus *et al.*⁴ and Morgan *et al.*⁵ look at the effects of immunization with β -amyloid peptide on learning and memory, as well as on brain damage, in mouse models^{8,9}. Their results support the previous observation³ of a reduction in the formation of amyloid deposits. But they go further, to show that immunization also offers the mice some protection from the 'spatial' learning deficits that normally accompany plaque formation.

Alzheimer's disease is characterized by both dementia — the near complete loss of mental function — and the abnormal physical changes that occur in brain tissue. But it has been difficult to determine which, if any, of these pathological features leads to dementia. One of the two major competing theories suggests that dementia is caused by the abnormal accumulation of β -amyloid peptide,

Box 1 Creating mouse models of human disease

Until the genetic mutations in humans that lead to Alzheimer's disease were discovered, the creation of a realistic animal model seemed unattainable. Although many of the key pathological features of the disease were known, there was no reliable way of reproducing them in a mouse. But once it became clear that mutations in the amyloid precursor protein could cause Alzheimer's disease in humans, the way seemed clear (as it did for animal models of Huntington's disease following a similar discovery).

The strategy is to introduce a mutant copy of the relevant gene into mice by injecting it into a fertilized mouse egg, and then breeding and testing the resulting transgenic mouse. In the case of Alzheimer's disease, the goal was to



produce an animal that had all the pathological hallmarks of human Alzheimer's disease. In particular, the mice should show age-related changes in behaviour, as well as deposition of β -amyloid into insoluble plaques (pictured left), the formation of neurofibrillary tangles within neurons, and the loss of neurons throughout the forebrain. (The right-hand image shows the scarcity of plaques in this mouse model

after vaccination with the β -amyloid peptide.) Although many of these features have been reproduced in one or more lines of transgenic mice, no one model has captured the full range of pathology. This might be seen as a rather inexplicable setback. But it can also serve as an opportunity to understand the contributions of individual pathological features to cognitive loss.

P.F.C.

deposited outside brain cells either in essentially insoluble 'plaques' (dense accumulations of proteins with β -amyloid peptide at their core), or in a more soluble form. If this 'amyloid' hypothesis is correct, then — theoretically — Alzheimer's disease could be treated by preventing β -amyloid peptide from forming or accumulating. The competing hypothesis is that a different protein, called tau, is the main cause of Alzheimer's dementia (or that amyloid and tau are equally culpable). If this is true, treatments aimed at the β -amyloid peptide might not be effective¹⁰.

So it will be important to have benchmarks in addition to the deposition of β -amyloid peptide for testing the effectiveness of treatments. For example, if any procedure reliably affects both β -amyloid accumulation and cognitive function in a mouse model, it will provide support for the amyloid hypothesis. It will also inspire greater confidence in that procedure's potential for treating humans.

Behavioural deficits in the mouse models used in the three new papers^{4–6} coincide with

higher concentrations of the β -amyloid peptide in the brain, or with the deposition of this peptide into plaques. But although failure to observe this correlation would have damaged the amyloid hypothesis, on its own the correlation does not provide proof of the theory. So it is significant that Janus *et al.*⁴ and Morgan *et al.*⁵ both found that simply hampering the formation of amyloid plaques is sufficient to prevent the decline of learning and memory.

Each group used one test of spatial memory, in which mice had to swim to and mount a platform located invisibly beneath the surface in a pool of water. Morgan *et al.* tested 'episodic-like' memory: the platform remained in the same location for every trial in one day, but was moved about randomly from day to day. Here, the mice cannot use their long-term memory of the location of the platform on previous days, but must rely on their short-term memory of its location on the day in question. Janus *et al.*, meanwhile, tested 'spatial-reference' memory: the platform remained in the same location during the five

days of testing. They also used a 'longitudinal' experimental design in which they tested the same mice every four weeks for roughly three months. They could therefore track changes in performance over time — much as would be done in human clinical studies. Given these differences, it is remarkable that both groups find that immunization with β -amyloid peptide offered significant protection from the age- and amyloid-dependent performance deficits seen in non-immunized controls.

Janus *et al.*⁴ also show that when the mice are immunized with the β -amyloid peptide in the so-called ' β -pleated sheet' form that will eventually be deposited, they produce antibodies that seem to be specific to this form. The result is a significant reduction in plaque formation and the number of amyloid 'fibrils', but there is no great difference in the overall level of β -amyloid peptide. This suggests that β -amyloid in its more soluble form does not contribute to Alzheimer's dementia. But, although Morgan *et al.*⁵ used the more soluble form of β -amyloid peptide for vaccination, they find modest but significant reductions in the proportion of the brain's neocortical and hippocampal regions that is covered by amyloid plaques. Both groups suggest that either a small or a selective reduction in β -amyloid deposition may be sufficient to protect against dementia.

The symptoms of Alzheimer's disease worsen with age, and so a valid animal model must also be age dependent. In the model used by Chen *et al.*⁶, there is a correlation between age (and β -amyloid deposition) and the impairment of some forms of learning, such as episodic-like memory. But some other learning defects in these mice are independent of both age and β -amyloid deposition. For example, some aspects of the episodic-like task are impaired in mice that are too young to show plaques, and object-recognition memory seems unaffected at any age⁶. The tasks used by Morgan *et al.* appear to be age related, but it is more difficult to assess those of Janus *et al.* because the mice were retested over several months. Further investigations of correlations among different behavioural measures and different pathological features will be important to strengthen the validity of the mouse models.

All in all, though, these three papers^{4–6} give cause for optimism. But the range of behaviours tested was rather narrow, and the mechanism of action of the β -amyloid vaccine is not yet fully understood. Other possible treatments, such as anti-inflammatory drugs¹¹ and inhibitors of enzymes that produce the β -amyloid peptide from its precursor protein^{12,13}, should be developed further and tested in these and other behavioural tasks, as should markers of other changes that accompany ageing and amyloid accumulation in mice¹⁴. The way will then be well prepared for clinical tests of treatments for Alzheimer's disease. ■

Paul F. Chapman is at the Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3US, UK.

e-mail: chapmanpf@cf.ac.uk

1. Bornemann, K. D. & Staufenbiel, M. *Ann. NY Acad. Sci.* **908**, 260–266 (2000).
2. van Leuven, F. *Prog. Neurobiol.* **61**, 305–312 (2000).
3. Schenk, D. *et al. Nature* **400**, 173–177 (1999).
4. Janus, C. *et al. Nature* **408**, 979–982 (2000).
5. Morgan, D. *et al. Nature* **408**, 982–985 (2000).

6. Chen, G. *et al. Nature* **408**, 975–979 (2000).
7. Games, D. *et al. Nature* **373**, 523–527 (1995).
8. Holcomb, L. *et al. Nature Med.* **4**, 97–100 (1998).
9. Hsiao, K. *et al. Science* **274**, 99–102 (1996).
10. Wischik, C. M., Edwards, P. C. & Harrington, C. R. in *Neurobiology of Alzheimer's Disease* (eds Dawbarn, D. & Allen, S. J.) 89–148 (BIOS Scientific, Oxford, 1995).
11. Lim, G. P. *et al. J. Neurosci.* **20**, 5709–5714 (2000).
12. Li, Y. M. *et al. Nature* **405**, 689–694 (2000).
13. Vassar, R. *et al. Science* **286**, 735–741 (1999).
14. Chapman, P. F. *et al. Nature Neurosci.* **2**, 271–276 (1999).

Astronomy

The Big Bang is bang on

John Bahcall

Did the Universe really start in a hot Big Bang? New measurements of the temperature of the Universe when it was young provide exciting confirmation that it was indeed hotter in the past.

The Universe is filled with unimagined things of great beauty and enormous significance, just waiting to be discovered. On page 931 of this issue, for example, Srianand, Petitjean and Ledoux¹ report the discovery of a rare set of features (absorption lines) in the light from a distant quasar. These features make it possible to test rigorously the fundamental idea, underlying all of modern cosmology, that the Universe was hotter when it was younger. The results reported by Srianand and colleagues are a convincing test of this idea and represent a landmark result.

Two twentieth-century discoveries have shaped the way humans view their place in the Universe. First, Edwin Hubble² discovered that the Universe is expanding — all other galaxies are moving away from us at a speed that is approximately proportional to their distance from our Galaxy. Second, in 1965, Arno Penzias and Robert Wilson discovered³ that the local part of the Universe is permeated with microwave radiation, called the cosmic microwave background (CMB).

Nearly all physicists and astronomers interpret the CMB as the leftover glow from a time when the Universe was very hot and very small. According to the popular Big Bang theory⁴, the Universe expanded from a tiny volume to its present vast expanse over the past 10^{10} years. As the initial gas of photons expanded, it cooled until the average wavelength of the cosmic radiation fell into the microwave region that we observe today. This theory predicts that the CMB temperature is the same everywhere at a given universal time and that the temperature increases in a simple and specific way as we look back to earlier and earlier times.

But is this really the story of our Universe? Did it really go from hot to cold? All of the widely discussed recent measurements^{5,6} of the CMB have been made locally, on properties of the microwave radiation near the Earth. Astronomers and physicists are opti-

mistic that they will be able to use these local measurements to determine the parameters that characterize the history of the Universe. They hope, for example, to determine whether the Universe will expand forever and to find out how much matter it contains.

In the late nineteenth century, physicists had a similarly grand concept of the cosmos. They supposed that the Universe was filled with an unseen quantity that they called the 'ether', a hypothesis that bears some resemblance to present-day ideas of dark matter. All radiation was supposed to propagate in the ether. But when Michelson and Morley searched experimentally for the ether, they couldn't find it. Physics had to be revised and the result was Einstein's special theory of relativity, which revolutionized our ideas about space and time. Would something similar happen when astronomers searched for data on the CMB in distant clouds?

Srianand and colleagues¹ analyse light from a distant quasar, one of the most luminous objects in the Universe. The light they studied was absorbed by carbon atoms and hydrogen molecules in a remote cloud of gas and dust when the Universe was only about one-sixth of its present age. Fortunately, the absorbed light provides several clues to what was happening at this early time. The astronomers used the ultraviolet and visible spectrograph (UVES) on the European Southern Observatory's 8.2-metre telescope at the Paranal observatory in Chile, and have obtained some wonderful results.

The cloud discovered by Srianand *et al.* is unique among the many hundreds of thousands of distant clouds whose absorption of light has been analysed so far. It is almost as if nature planted an abundance of clues in this anonymous cloud in order to allow some lucky researchers to infer the temperature of the CMB when the Universe was young. The absorption of quasar light by neutral and once-ionized carbon atoms shows that some of the cloud's atoms are in atomic states

called 'fine-structure states', which are only slightly more energetic than the lowest energy states of carbon. The difference in energy between a typical fine-structure state and the lowest energy state of carbon is comparable to the CMB temperature predicted by Big Bang theory in the vicinity of the distant cloud.

The different atoms and molecules that absorbed light in this remote cloud enable the astronomers to demonstrate that there is only one reasonable explanation for the number of atoms that are observed to be in fine-structure states. Srianand *et al.* conclude that there must be an ambient radiation field around the cloud corresponding to a temperature of between 6 K and 14 K. Bingo! This value agrees with the temperature of 9 K predicted by the Big Bang theory to be the CMB temperature at the position of the cloud. This is three times as high as its temperature near Earth (2.7 K), as expected.

The idea of using atomic fine-structure lines to test Big Bang cosmology was first suggested⁷ in 1968, only three years after the discovery of the CMB. Over the past three decades, a series of ever more precise measurements, or upper-limit determinations of the CMB temperature in distant clouds, were made using fine-structure lines⁸. Generations of sensitive telescopes and spectroscopic instruments had to be developed before the latest robust test was possible. It also took good luck to find a cloud that contained the necessary clues — carbon atoms in fine-structure states, hydrogen molecules in various excitation states, and light absorption by iron, silicon and magnesium atoms.

In future, astronomers will be able to use the latest 8- and 10-metre telescopes to look for other rare absorbing clouds at different distances from us (that is, at different ages of the Universe) and in different directions (to check that the CMB is the same in every direction). Tests of the Big Bang theory using fine-structure lines will become more precise and more numerous in the next few years as larger telescopes and better instrumentation become widely available.

The Big Bang theory has survived a crucial test. The theory would have been abandoned if astronomers had found that clouds at earlier times had lower temperatures than predicted. So while the result reported by Srianand and colleagues is a tremendous achievement, I confess to feeling a little disappointed, and I am sure that some of my more rebellious colleagues will secretly feel the same way. I am happy that the Big Bang theory passed this test, but it would have been more exciting if the theory had failed and we had to start looking for a new model of the evolution of the Universe. ■

John Bahcall is at the Institute for Advanced Study, Natural Sciences, Einstein Drive, Princeton, New Jersey 08540, USA.

e-mail: jnb@ias.edu

1. Srianand, R., Petitjean, P. & Ledoux, C. *Nature* **408**, 931–935 (2000).
2. Hubble, E. A. *Proc. Natl Acad. Sci. USA* **15**, 168–173 (1929).
3. Penzias, A. A. & Wilson, R. *Astrophys. J.* **142**, 419–421 (1965).
4. Weinberg, S. *The First Three Minutes: A Modern View of the*

Origin of the Universe (Basic Books, New York, 1977).

5. De Bernardis, P. *et al.* *Nature* **404**, 955–959 (2000).
6. Hanany, S. *et al.* <http://xxx.lanl.gov/astro-ph/0005123>
7. Bahcall, J. N. & Wolf, R. W. *Astrophys. J.* **152**, 701–729 (1968).
8. Ge, J., Bechtold, J. & Black, J. H. *Astrophys. J.* **474**, 67–73 (1997).

Evolutionary biology

Cooperation can be dangerous

Richard H. Kessin

In social situations, opportunities arise for some individuals to take advantage of others. This happens in wild populations of the social amoeba *Dictyostelium discoideum*.

Most of the organisms studied by developmental biologists — the fruitfly *Drosophila*, for instance — arise from a fertilized egg. So, apart from the mature reproductive cells, the cells from which these organisms are made are genetically identical. No cell has a genetic advantage over the others, and any variants that do arise are excluded from the germ line. Because of the lack of genetic diversity among their cells and the

presence of a germ line, cheating by individual cells is restricted in most multicellular organisms.

But there is another route to multicellularity, taken by organisms such as *Dictyostelium* and *Myxococcus*, a group of social bacteria. As they describe on page 965 of this issue¹, Strassmann and colleagues have studied the curious cooperative behaviour of *Dictyostelium*, and find that some genetic

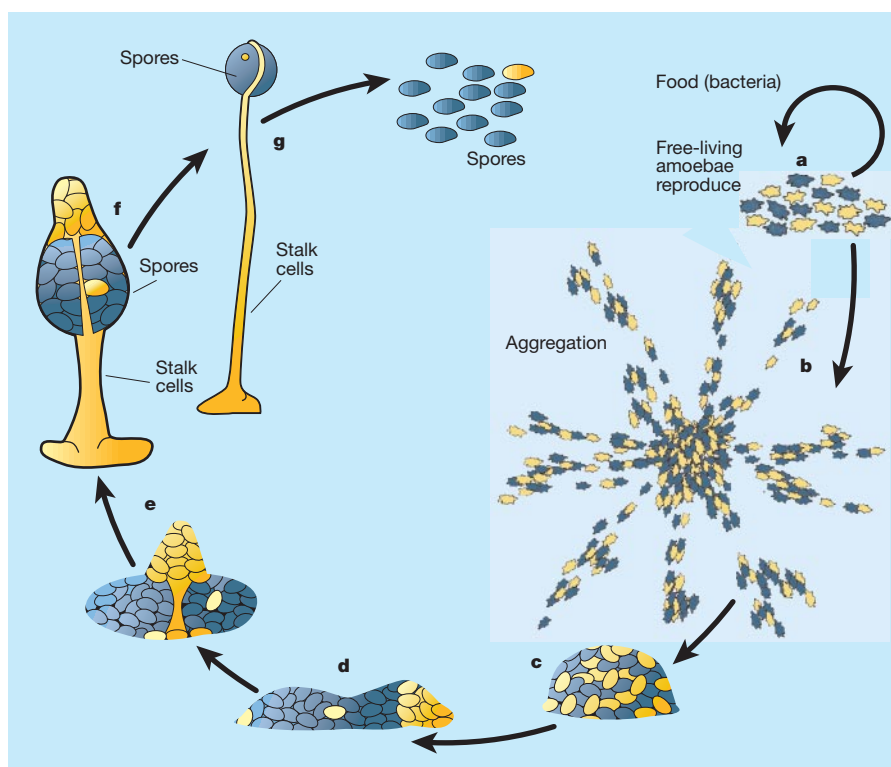


Figure 1 Cheaters in the *Dictyostelium* life cycle. a, *Dictyostelium* amoebae live off bacteria in the soil. b, When the bacteria are consumed and starvation is imminent, the amoebae stop dividing and activate several genes that allow them to aggregate by chemotaxis towards cyclic AMP diffusing from centrally located cells. These aggregates, which can contain up to 100,000 cells, transform into motile slugs (c, d) and finally into fruiting bodies (e–g). The fruiting bodies contain 80,000 viable spores supported by 20,000 dead stalk cells. If the cells of two natural isolates are mixed, they will aggregate. But in some combinations, as Strassmann *et al.*¹ show, one isolate predominates in the spore population. Here, equal numbers of yellow and blue cells aggregate, but eventually the blue cells dominate the spore mass. How this advantage is achieved is not known, but it could involve manipulation of the signals that control the 80:20 ratio of spore and stalk cells. For a dynamic view of *Dictyostelium* development see ref. 10. Graphic modified from ref. 11.

variants cheat on their partners during the process of spore formation (Fig. 1).

In times of plenty, *Dictyostelium* live as individual amoebae in soil, preying on bacteria. But when food runs out and starvation is imminent, the previously independent amoebae form aggregates of as many as 100,000 cells². These aggregates go on to form motile slugs, which can move to fresh pastures. The slugs then construct fruiting bodies in which 80% of the cells form viable spores and the other 20% die in making a stalk. It is the death of the stalk cells, which seemingly sacrifice themselves for the common good, that makes *Dictyostelium* interesting to evolutionary biologists.

Strassmann, Zhu and Queller¹ show that some isolates of *Dictyostelium* have an advantage over others in chimaeras — mixtures of genetically different amoebae. These cheater isolates make a minimal contribution to the stalk population, leaving that onerous function to their partners. The authors made mixtures of isolates collected from the same area, let them form slugs, and then sampled regions fated to be spore (the rear of the slug) or stalk (the front). Using the polymerase chain reaction to amplify certain DNA sequence differences between two isolates, they could ask whether a strain was making its appropriate contribution to the stalk population. Often, but not always, the answer was 'no'. In one combination, a participant contributed almost no cells to the stalk. Life in the soil is not fair.

Several *Dictyostelium* variants, natural and laboratory-induced by mutagenesis, that achieve an advantage in chimaeras have been described^{3–6}. But all cheaters isolated from these variants had defective development without a partner (they would not have been noticed otherwise). Such strains would have been selected against in the wild, where single *Dictyostelium* spores are often dispersed to form clonal colonies — colonies in which all members have an identical genetic constitution.

By contrast, the isolates used in the experiments of Strassmann *et al.* developed normally without a partner. All also formed chimaeras with each other — that is, they aggregated, indicating that no incompatibility mechanisms had evolved among them. The authors suggest a view of *Dictyostelium* as a society of competing, genetically distinct individuals, which strike a balance between competition and cooperation depending on their relatedness, and so have much in common with insect societies. As Strassmann *et al.* point out, parasitism in insect societies is often based on stealing chemical signals. It is possible that the cheater mutants of *Dictyostelium* likewise send a signal to the other members of the aggregation that makes them form the stalk.

Evolutionary and developmental biologists will view these results differently. Evolu-

tionists will wonder why certain variants don't eliminate the others and become dominant. And if isolation is the way *Dictyostelium* avoids competition, how did the genetic diversity arise that is a prerequisite for the evolution of this elaborate fruiting behaviour? Is *Dictyostelium* evolution a story of constant warfare, of different strains angling for advantage and counter-advantage? Are there mechanisms that limit the invasion of one strain by another?

For comparison, there are other organisms that display similar behaviour. *Myxococcus xanthus*, for instance, develops into spores after a process of aggregation, and some mutant strains take advantage of others during fruiting⁷. Another example is *Botryllus schlosseri*, which is a relatively complex organism with elaborate feeding and circulatory systems. Individuals attach to seaweed and, when adjacent to each other, closely related individuals fuse, whereas more distantly related ones undergo an incompatibility reaction⁸. In some fusions, the germ cells of one variant invade the cytoplasm of a neighbouring variant and produce descendants at its expense.

As for developmental biologists, they have not yet succeeded in explaining how a properly proportioned *Dictyostelium* fruiting body forms. They can ask whether the approach used by Strassmann *et al.* will provide a way of identifying the signals that the developing

amoebae use to regulate their proportions and position. Certain mutagenesis schemes allow the affected genes to be recovered from cheater mutants^{6,9}. Only one such gene has been recovered so far and it dramatically affects the ratio of stalk cells to spore cells. It remains to be seen whether it proves to be the useful end of a string on which one can pull to haul in other genes that rule the elegant cell proportionality of the fruiting bodies. Given the molecular genetic and genomic opportunities offered by *Dictyostelium*, a combination of developmental and evolutionary approaches may be the most promising way to tackle these questions. ■

Richard H. Kessin is in the Department of Anatomy and Cell Biology, Columbia University, 630 West 168th Street, New York, New York 10032, USA.
e-mail: rhk2@columbia.edu

1. Strassmann, J. E., Zhu, Y. & Queller, D. C. *Nature* **408**, 965–967 (2000).
2. Kessin, R. H. *Dictyostelium — Evolution, Cell Biology, and the Development of Multicellularity* (Cambridge Univ. Press, January 2001).
3. Filosa, M. F. *Am. Nat.* **XCVI** (no. 887), 79–92 (1962).
4. Buss, L. W. *Proc. Natl Acad. Sci. USA* **79**, 5337–5341 (1982).
5. Dao, D. N., Kessin, R. H. & Ennis, H. L. *Microbiology* **146**, 1505–1512 (2000).
6. Ennis, H. L., Dao, D. N., Pukatzki, S. U. & Kessin, R. H. *Proc. Natl Acad. Sci. USA* **97**, 3292–3297 (2000).
7. Velicer, G., Kroos, L. & Lenski, R. E. *Nature* **404**, 598–601 (2000).
8. Stoner, D. S., Rinkevich, B. & Weissman, I. L. *Proc. Natl Acad. Sci. USA* **96**, 9148–9153 (1999).
9. Nelson, M. K. *et al. Dev. Biol.* **224**, 42–59 (2000).
10. <http://dicty.cmb.nyu.edu/dicty/pnd/>
11. Kessin, R. H. & van Lookeren Campagne, M. M. *Am. Sci.* **80**, 556–565 (1992).

Fundamental constants

Measuring big *G*

Terry Quinn

Newton's constant, *G*, which governs the strength of the gravitational attraction between two masses, is difficult to measure accurately. A new set of experiments aims to end 200 years of uncertainty.

Newton's gravitational constant, *G*, differs from all the other fundamental constants of physics in that there is no complete theory that links gravity to the other forces of nature. Hence there is no definitive relationship between *G* and the other fundamental constants. It also stands apart in that the accepted uncertainty of 0.15% for *G* is many orders of magnitude larger than for the other fundamental constants (as given in the 1998 CODATA report¹). It is astonishing that 200 years after the famous experiment of Henry Cavendish to 'weigh the Earth' we seem to have improved on his result by only a factor of ten. Now, in *Physical Review Letters*, Gundlach and Merkowitz² report a value for *G* with an uncertainty of about 0.001% — so is *G* finally becoming respectable?

The measurement of *G* is simple in theory: take two spherical masses *M*₁ and *M*₂ at a known distance *r* apart, and measure the

gravitational attractive force between them, $F = GM_1M_2/r^2$. The problem is that the gravitational attraction between any two laboratory-sized masses is simply too small to measure accurately. For example, if we take two 1-kg masses, 10 cm apart, the gravitational attraction between them is about 6×10^{-9} N. Put another way, if one of the masses is fixed, the acceleration of the other towards it is 6×10^{-9} m s⁻². Such an acceleration is infinitesimal compared with the local acceleration due to Earth's gravity, *g*, of 9.8 m s⁻².

From the time of Cavendish, the preferred method of measuring *G* has been the torsion balance, in which the restoring force of a twisted fibre balances the weak gravitational torque (twisting force) produced by the attraction between several test masses and a pendulum suspended at the end of the fibre (Fig. 1, overleaf). This method provides an excellent way of decoupling the gravitational



100 YEARS AGO

A few weeks ago the new anthropological collections in the American Museum of Natural History in New York were opened to the public, and these valuable collections now occupy five halls, and others are being provided. We learn from our contemporary, *Science*, that the accessions to the anthropological collections of the museum obtained during the last three years have largely been due to extended scientific research undertaken by the institution... an endeavour has been made to build up representative collections and to obtain, at the same time, the fullest and most detailed information in regard to specimens, so that each addition to the exhibit of the museum can be made thoroughly instructive and will represent a material contribution to science. There is no doubt this is the best way to build up a museum, and it is to be deplored that various museums of the British Islands do not follow the example so worthily set by this and other American museums. Our English method is rather to wait like a spider in its web in the hope that something will eventually be caught ... we are content with occasional specimens which usually have no history, or at most a very imperfect one, and for these we often have to pay a stiff profit to a dealer.

From *Nature* 20 December 1900.

50 YEARS AGO

For one of the staff — one of the “working staff” as Sir James Dewar used to call us — to be asked to give a Friday Evening Discourse is, I think, quite without precedent and I am very conscious of the honour the Managers have conferred on me in inviting me to give a talk about fifty years in the service of the Royal Institution. According to tradition, on one occasion many years ago, a Friday Evening lecturer had a sudden attack of stage fright at the last moment and, being unattended, fled. Fortunately, Faraday was present and stepped into the breach. So that there should be no recurrence of a similar catastrophe, every Friday evening someone waits outside with the lecturer to ensure that he enters this theatre as the clock strikes. For many years that has been one of my duties; but to-night the position has been reversed and I have been the guarded one. I can assure you I have every sympathy with Wheatstone, the one that ran away. *Ralph Cory, Librarian of the Institution.*

From *Nature* 23 December 1950.

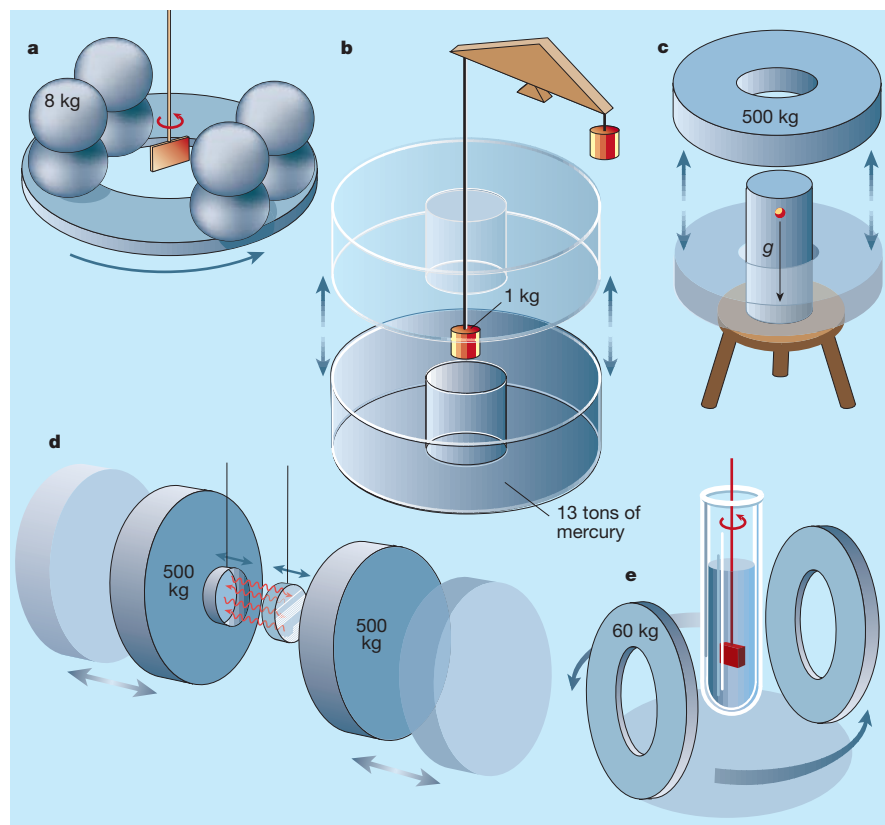


Figure 1 Updating Cavendish in measuring G . a, Gundlach and Merkowitz² have set a new standard for measuring G with a torsion-balance experiment in which eight spheres on a rotating disc turn to follow a thin plate suspended on a fibre, driven by the gravitational torque from the spheres. There are other ways to measure G , for example: b, an experiment in which a beam balance is used to measure the weight of a 1-kg mass on the pan of the balance when 13 tons of mercury are displaced from above to below it; c, an experiment in which laser interferometry is used to measure the change in downward acceleration of a falling body when a 500-kg mass is displaced from above to below it; d, an experiment using the gravitational attraction of large 500-kg masses to displace small masses hanging from a pair of pendulums that act as optical or microwave cavities; and e, a cryogenic torsion balance in liquid helium in which doughnut-shaped masses (at room temperature) turn around a thin plate suspended from the cold fibre.

attraction from the effects of g . However, whether G is measured using a torsion balance or any other device, it is necessary to construct test masses whose dimensions and density are known with sufficient accuracy. If spherical or cylindrical masses are used that have perfect geometry, the effects of density variation can be eliminated by random changes in orientation. But this does not work if the geometry is not perfect and, in any case, becomes more difficult with larger masses.

In the absence of any advances in physics linking gravity to the rest of science, there have been no really new methods of measuring G since the time of Cavendish. Despite a flurry of excitement over the ‘fifth force’ in the 1980s, or apparently strange gravitational effects acting on spinning rotors reported in the 1990s, Newton continues to reign supreme in laboratory gravitation. Nevertheless, measurements of G hold great interest for both cosmology and particle physics; in the latter case it has been suggested that the compact dimensions predicted by ‘superstring theory’ might show up in the behav-

iour of G at small (< 1 mm) distances³.

The current interest in measuring G was stimulated by the publication⁴ in 1996 of a value for G that differed by 0.6% from the accepted value given in the previous 1986 CODATA report. To take account of this, the 1998 CODATA report recommends a value for G of $6.673 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$ with an uncertainty of 0.15%, some ten times worse than in 1986. Whereas the other fundamental constants were more accurately known in 1998 than in 1986, the uncertainty in G increased dramatically. The G community appeared to be going backwards rather than forwards.

Since 1998, several groups around the world have set about measuring G , using a range of different methods. At a symposium held in London in 1998 to celebrate the bicentenary of Cavendish’s experiment, reports of eight experiments then under way were presented⁵, some of which are shown in Fig. 1. The target uncertainty for these experiments is between 0.01% and 0.001%. Mostly preliminary results have been published so

far, but all, including the new result of Gundlach and Merkowitz², are in rough agreement and do not support the controversial 1996 value. This is good news because the different techniques have different sources of error, reducing the likelihood of systematic errors affecting the final result.

But not all the improvements in accuracy have resulted from new techniques — there have also been similar advances in torsion balance experiments^{6–9}. Gundlach and Merkowitz's result is based on a torsion-balance method that has several advantages over previous experiments. The design eliminates much of the uncertainty resulting from the distribution of the pendulum mass suspended from the torsion fibre (Fig. 1a). If this suspended mass is a thin flat plate, corrections due to the shape, the total mass and the mass distribution of the plate cancel out in the analysis. Using a set of eight spherical test masses, almost all effects resulting from the suspended mass become negligible.

Gundlach and Merkowitz measure the angular acceleration of their suspended mass by rotating the whole system on a turntable. This has the advantage that the torsion wire is never twisted, so the measurement of G is independent of anelasticity in the fibre properties¹⁰. The rotation also averages out the effects of local gravity gradients. The result has an uncertainty of 0.001%, well below anything claimed before. But their value for G is 0.024% above the original 1986 CODATA value. Is it right?

The only way we can generate confidence in a result is by having several independent measurements made using different experimental techniques. Agreement has yet to be achieved at the 0.01% level, let alone 0.001%. Although we do not yet have a theoretical prediction for G against which to test the results, a reliable and accurate experimental value is important for testing future theories. The G community awaits with great interest the final results of the other measurements in progress. In the meantime, the result of Gundlach and Merkowitz sets a new standard for all G experiments to match. ■

Terry Quinn is at the Bureau International des Poids et Mesures, Pavillon de Breteuil, F-92312 Sèvres, France.

e-mail: tquinn@bipm.org

- Mohr, P. J. & Taylor, B. N. *J. Phys. Chem. Ref. Data* **28**, 1713–1852 (1999); <http://physics.nist.gov/cuu/Constants/index.html>
- Gundlach, J. H. & Merkowitz, S. M. *Phys. Rev. Lett.* **85**, 2869–2872 (2000).
- Arkani-Hamed, N. *et al. Phys. Lett. B* **429**, 263–272 (1998).
- Michaelis, W., Haars, H. & Augustin, R. *Metrologia* **32**, 267–276 (1996).
- Quinn, T. J. & Speake, C. C. *Meas. Sci. Technol.* **10**, 421–530 (1999).
- Quinn, T. J., Speake, C. C. & Brown, L. M. *Phil. Mag. A* **65**, 261–276 (1992).
- Kuroda, K. *Phys. Rev. Lett.* **75**, 2796–2798 (1995).
- Fitzgerald, M. P. & Armstrong T. R. *IEEE Trans. Instrum. Meas.* **44**, 494–497 (1995).
- Richman, S. J., Quinn, T. J., Speake, C. C. & Davis, R. S. *Meas. Sci. Technol.* **10**, 460–466 (1999).
- Rose, R. D. *et al. Phys. Rev. Lett.* **23**, 655–658 (1969).

Neurobiology

Moving forward by looking away

Larry Snyder

In one model of the brain, a central processing region is sandwiched between separate input and output areas. But studies of humans, and now monkeys, hint that this model may be too simplistic.

A conventional approach to studying the brain is to determine whether a particular signal or region is mainly sensory or motor in nature, reflecting what comes into the brain or what flows out, respectively. In this view, the input region (the sensory cortex) processes sensory signals to generate several neuronal representations of the information. Next, a central processor, found outside the sensory cortex, uses information about the goals of the task in hand to select the most appropriate sensory representation for directing motor output. This selection is the key step in the 'sensorimotor transformation'. But in another model, task-specific information influences the processing of incoming signals from an early stage. Pure sensory signals are rare, and the sensorimotor transformation is distributed throughout the brain. This view has begun to take hold among researchers who study the human brain¹. But convincing evidence is still needed. On page 971 of this issue, Zhang and Barash² provide just such evidence from studies of monkeys.

The task of moving one's eyes to a visual target involves a clear sensory input (from the target) and motor output (the eye movement), so it seems ideal for studying sensorimotor transformations. But experiments based on eye movements are often handicapped by a difficulty in distinguishing neuronal representations of what we see and what we plan to do. Imagine being inspired by the sight of a travel poster and deciding to visit London. Before you choose how and when to travel, your plan is best represented by the image on the poster — the poster is both the stimulus that evokes your plan, and an early representation of the plan itself.

A similar confusion occurs when studying directed eye movements. A simple way to introduce a distinction between representations of visual stimuli and plans to move the eyes to those stimuli is to study both prosaccades, which are eye movements towards a visual target, and antisaccades, eye movements away from the target³. The colour of a 'fixation spot' tells the subject whether to move their eyes towards or away from the target. This combination allows the researcher to determine whether neuronal activity is more closely related to the stimulus or the movement goal, and therefore where task-specific processing begins.

The lateral intraparietal area (LIP) in the posterior parietal cortex of monkey brains is likely to participate in, or even initiate,

sensorimotor transformations during saccades⁴. Neurons in this region link sensory and motor areas. Individual neurons are activated when a behaviourally relevant stimulus appears within the cell's receptive field. LIP neurons also fire shortly before an eye movement to a remembered or visible target in the receptive field.

Last year, Gottlieb and Goldberg⁵ recorded the activity of single neurons in the LIP while monkeys performed prosaccades and antisaccades. They found that most of the sampled neurons represented the location of the visual target. Few represented the direction of the eye movement; when they did, they fired quite late, around the time of the movement.

Zhang and Barash² have now repeated Gottlieb and Goldberg's experiment with a slight change, and obtained very different results. Rather than having the monkeys move their eyes as soon as they saw a target, Zhang and Barash trained the animals to delay their movements. In this way, signals time-locked to the sensory stimulus could be more easily distinguished from signals related to the motor response. Neuronal activity began hundreds of milliseconds before the signal to make the saccade and well after the sensory stimulus, reflecting the direction of the movement, not the location of the target. So, the sensorimotor transformation occurs in the LIP long after the stimulus appears but well before the eye movement begins.

Zhang and Barash also report evidence of an earlier sensorimotor transformation in the LIP. During both prosaccade and antisaccade trials, many LIP neurons become active shortly after a stimulus falls inside their receptive field. The authors show that a subset of these cells also becomes active soon after a stimulus appears outside their receptive field that will direct an antisaccade so that the stimulus falls into the field (Fig. 4b, page 973). Zhang and Barash refer to this as 'paradoxical activity'. These neurons show activity that is time-locked to the onset of the visual stimulus, suggesting that they are driven directly by visual inputs. But they also respond to a stimulus outside the receptive field on an antisaccade trial, consistent with an input that codes the end of the planned movement rather than the target's location. This activity probably represents the start of the sensorimotor transformation for an antisaccade.

But this transformation is incomplete on at least two counts. First, these neurons often fire when a stimulus outside their receptive

field directs an antisaccade back into that field. Yet when a stimulus appears inside the receptive field and directs a saccade to the outside, there is no complementary reduction in activity. Second, activity that encodes the upcoming direction of a saccade would be expected to persist until the movement occurs. But the paradoxical activity subsided long before the late activity began to build up. The fact that the representation of antisaccades in the LIP is incomplete is consistent with this sensorimotor transformation being spread out over several regions of the brain.

Why did Zhang and Barash², using delayed saccades, observe a partial early and a complete late sensorimotor transformation in LIP, whereas Gottlieb and Goldberg⁵ observed only sensory representations and rare motor responses during saccades? Antisaccades began in Gottlieb and Goldberg's task at a similar time to the paradoxical activity in the delayed task. Perhaps this activity was also present in Gottlieb and Goldberg's task but was obscured by the motor response, which occurred at about the same time. By delaying the saccade, Zhang and Barash's set-up may have allowed further time for both early (paradoxical) and late transformation signals to build up in LIP, increasing the chances that they would be observed. Or perhaps the insertion of a delay fundamentally changed the way in which the

animals solved the task. It would be worthwhile repeating the two experiments in the same animal and the same cells.

Nonetheless, Zhang and Barash's data² clearly show that, in a practised, delayed antisaccade task, the sensorimotor transformation is seen — and perhaps begins — in the LIP. The authors argue that the timing of the paradoxical activity preceding antisaccades suggests that it is produced by visual inputs (albeit non-standard inputs). They compare this to a report of a rerouting of visual inputs into the LIP⁶. Yet what Zhang and Barash describe is no mere rerouting of the sensory information. To decide whether to perform a prosaccade or an antisaccade, the animals must modify their behaviour on a trial-by-trial basis, requiring both sensory and non-sensory information. Fortunately, these are experimentally tractable issues and the stage is now set to address them. ■

Larry Snyder is in the Department of Anatomy and Neurobiology, 660 South Euclid Avenue, St Louis, Missouri 63110, USA.

e-mail: larry@eye-hand.wustl.edu

1. Shulman, G. L. *et al.* *Cereb. Cortex* **7**, 193–206 (1997).
2. Zhang, M. & Barash, S. *Nature* **408**, 971–975 (2000).
3. Hallet, P. E. *Vision Res.* **18**, 1279–1296 (1978).
4. Gnadt, J. W. & Andersen, R. A. *Exp. Brain Res.* **70**, 216–220 (1988).
5. Gottlieb, J. & Goldberg, M. E. *Nature Neurosci.* **2**, 906–912 (1999).
6. Duhamel, J. R., Colby, C. L. & Goldberg, M. E. *Science* **255**, 90–92 (1992).

Magnetoelectronics

Teaching magnets new tricks

David D. Awschalom and Roland K. Kawakami

A way to control magnetism in semiconductors using an external electric field has been shown for the first time. This long-awaited result could lead to new types of information-storage devices.

Information technology is rooted in semiconductors, such as silicon, and ferromagnetic materials, such as iron. Information processing and computation are based on semiconductor transistors and integrated circuits, and information is stored magnetically on high-density hard drives. Even the use of high-speed optical communications begins and ends with semiconductor lasers and photodetectors.

The emerging field of 'spintronics' aims to combine ferromagnets with semiconductors to develop electronic devices that exploit the quantum-mechanical 'spin' of electrons as well as their charge¹. One aim is to integrate information storage with information processing, but a broader goal is to develop new functionality that does not exist separately in a ferromagnet or a semiconductor. To this end, investigators are searching for 'emergent behaviour' in combined ferromagnet/semiconductor structures, in which the whole is more than a sum of its parts.

On page 944 of this issue, Ohno *et al.*² provide a striking example of such behaviour. They show that the magnetic phase of a ferromagnetic semiconductor can be controlled using electric fields. In effect, Ohno *et al.* have created an electric 'switch' that

enables a permanent magnet to be turned on and off reversibly, using standard electronic techniques. Such behaviour has never been seen before in a ferromagnet. It works because the concentration of 'charge carriers' in the semiconductor (the negatively charged electrons and positively charged 'holes') can be altered with electric fields.

The material used in this study is a ferromagnetic semiconductor alloy known as indium manganese arsenide^{3–5}, (In,Mn)As. The typical concentration of Mn is less than 10%, and the Mn atoms replace some of the In atoms on the InAs lattice. The Mn atoms serve a dual role in determining the properties of the material. First, each Mn atom has a magnetic dipole moment (a 'north' and 'south' pole) and so acts as a small permanent magnet. Second, the Mn atoms are 'electron acceptors' which remove electrons from the material and so generate holes. These holes are the 'absence' of electrons, and so behave as positively charged particles. This is similar to how an underwater bubble is the 'absence' of water and behaves as an object with negative mass (underwater bubbles rise).

If all the magnetic moments of the Mn atoms are aligned, they add up to create a macroscopic magnetic dipole (as in a permanent bar magnet) and the material is in the 'ferromagnetic phase'. If the magnetic moments are randomly oriented, then the overall magnetic dipole strength is zero and the material is in the 'paramagnetic phase'. Under normal conditions, the material becomes ferromagnetic when it is cooled below a critical ordering temperature (T_c), which for (In,Mn)As is about 30 K.

What causes the Mn magnetic moments to align is not fully understood. But it is widely believed that the presence of holes is crucial for the process. Unlike the Mn atoms, which remain fixed in space, the holes can move throughout the material and act as go-betweens, coupling the behaviour of the spatially separated Mn atoms. Experimentally, this idea is supported by the observation that the value of T_c in fact varies with the hole concentration⁶ (see Fig. 1). Indeed, shining light

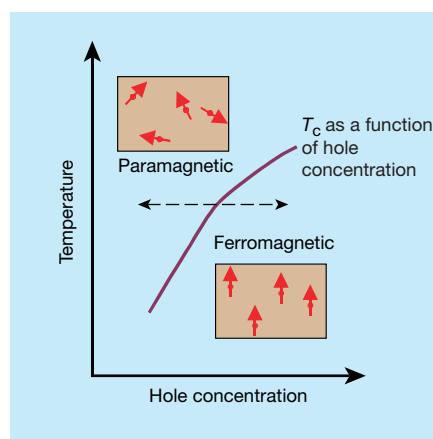


Figure 1 Behaviour of a magnetic semiconductor as a function of temperature and hole concentration (holes are positive charges). The orientation of the magnetic dipole moments of the manganese atoms (red arrows) in the indium manganese arsenide material (grey boxes) can be altered by a change in temperature — specifically by crossing the critical temperature, T_c . But the magnetic behaviour of the semiconductor can also be altered by changing the hole concentration (dashed line with arrows). In this way, Ohno *et al.*² drive their system from a ferromagnetic phase (aligned manganese moments) to a paramagnetic phase (randomly oriented manganese moments), and back, while the material remains at a fixed temperature.

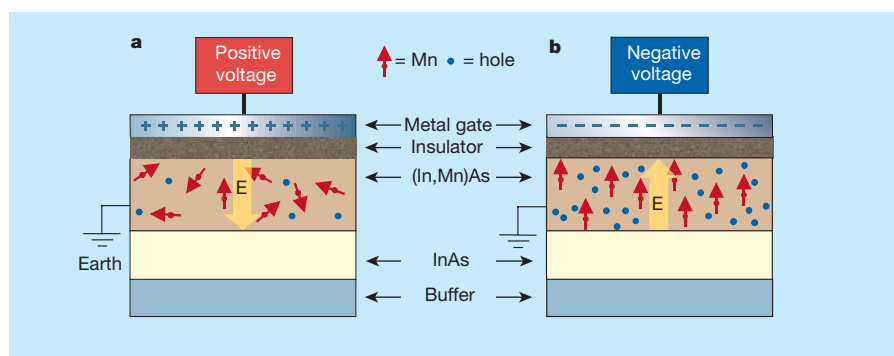


Figure 2 An electrically controlled magnetic switch. The device designed by Ohno *et al.*² has a layer of indium manganese arsenide, (In,Mn)As, material sandwiched within a field-effect transistor. **a**, A positive voltage on the metallic 'gate' electrode creates electric fields (orange arrow) that repel holes (blue circles), causing the Mn magnetic moments (red arrows) to orient randomly. **b**, A negative voltage on the gate electrode creates electric fields that attract holes, causing the Mn magnetic moments to align.

onto (In,Mn)As semiconductors and related magnetic alloys — thereby varying the concentration of charge carriers — can also produce changes in magnetization^{7,8}. If it is possible to control the hole concentration in real-time using an electric field, then it should be possible to drive the system reversibly from a ferromagnetic phase (aligned Mn moments) to a paramagnetic phase (randomly oriented Mn moments) and vice versa. Crucially, for practical applications, these changes in magnetization would occur at a fixed temperature (dashed line in Fig. 1).

In their work, Ohno *et al.*² change the hole concentration of an (In,Mn)As thin film by placing it in the active region of a field-effect transistor (FET). The FET structure shown in Fig. 2 is created by growing a thin layer of (In,Mn)As on an InAs-terminated buffer layer, followed by subsequent deposition of an insulating layer and a metallic 'gate' electrode. Applying a positive voltage to the gate (Fig. 2a) creates electric fields that repel the positively charged holes (because like charges repel). The reduced hole concentration in the (In,Mn)As layer makes the system paramagnetic. Applying a negative voltage to the gate (Fig. 2b) creates electric fields that attract the positively charged holes (because opposite charges attract). This increases the hole concentration and makes the system ferromagnetic. So by simply applying a voltage to the gate electrode, the magnet can be reversibly turned on (ferromagnetic with a net magnetic dipole moment) and off (paramagnetic with zero magnetic moment).

The concept behind Ohno *et al.*'s device is quite elegant, and begs the question 'why didn't anyone think of this before?'. Of course, they did, and there have been many unsuccessful attempts to build such devices. For instance, it has proved very difficult to change the charge-carrier concentration in metals (and most ferromagnets are metallic). In addition, there has been no success with devices made using gallium manganese arsenide, (Ga,Mn)As — which has a T_C as high as 110 K

— because of their high intrinsic carrier concentration⁵. Most semiconductor materials break down under high electric fields, which limits the extent to which one can control the carrier concentration in an FET. In an effort to reduce the hole density, the ferromagnetic semiconductor layers were made thinner, but ferromagnetism in (Ga,Mn)As disappears below a thickness of around 5 nm. To circumvent these difficulties, Ohno *et al.* used a different material, (In,Mn)As.

This newly emergent behaviour is quite promising, but do not expect to see this 'magnet switch' on your desktop tomorrow.

Meteorology

Oscillating opinion

Heike Langenberg

Why, since around 1960, have winters in northern Europe tended to become milder and wetter? The meteorological conditions responsible came under discussion at a meeting last month.

The North Atlantic Oscillation has a big influence on weather and climate in the Atlantic region, and so has been termed El Niño's northern cousin. But as was clear at a conference* on the topic, the phenomenon is more complicated than its counterpart in the tropical Pacific — which itself is far from understood. Not surprisingly, then, even the question "What exactly is the North Atlantic Oscillation?" is controversial.

The North Atlantic Oscillation (NAO) has been discussed for almost 70 years¹, and is generally defined as the principal pattern of wintertime atmospheric-pressure variability at sea level over the North Atlantic. The most widely used index for the period since 1864 is the difference between normalized atmospheric pressures at Stykkisholmur, Iceland, and Lisbon, Portugal². In the

The effects are seen at low temperatures (about 25 K) and with the application of large voltages (± 125 volts). So many challenges need to be overcome before this effect can be used in a commercially viable device. Nonetheless, this experiment is a 'proof of concept' for the idea that the magnetic properties of ferromagnetic semiconductors can be controlled using standard electronic techniques. This finding, along with the discovery of new ways to control electronic spin — for example, creating longer spin lifetimes in semiconductors⁹ and spin-injection from magnetic semiconductors¹⁰ — paves the way for practical spintronics.

David D. Awschalom and Roland K. Kawakami are at the Center for Spintronics and Quantum Computation, and the Department of Physics, University of California, Santa Barbara, California 93106, USA.

e-mail: awsch@physics.ucsb.edu

1. Prinz, G. A. *Science* **282**, 1660–1663 (1998).
2. Ohno, H. *et al.* *Nature* **408**, 944–946 (2000).
3. Ohno, H., Munekata, H., Penney, T., von Molnár, S. & Chang, L. L. *Phys. Rev. Lett.* **68**, 2664–2667 (1992).
4. Munekata, H., Zaslavsky, A., Fumagalli, P. & Gambino, R. J. *Appl. Phys. Lett.* **63**, 2929–2931 (1993).
5. Ohno, H. *Science* **281**, 951–956 (1998).
6. Matsukura, F., Ohno, H., Shen, A. & Sugawara, Y. *Phys. Rev. B* **57**, R2037 (1998).
7. Koshihara, S. *et al.* *Phys. Rev. Lett.* **78**, 4617–4620 (1997).
8. Wojtowicz, T., Kolesnik, S., Miotkowski, I. & Furdyna, J. K. *Phys. Rev. Lett.* **70**, 2317–2320 (1993).
9. Kikkawa, J. M. & Awschalom, D. D. *Physics Today* **52**, 33–39 (1999).
10. Ball, P. *Nature* **404**, 918–920 (2000).

positive phase — that is, in years with high pressure differences (see Fig. 1) — a strong storm track reaches far north, into the Barents Sea. This results in storms, rain and mild temperatures in northern Europe, while temperatures in Greenland and Labrador are colder than average. In a winter with a negative NAO index, this pattern reverses, as there is a weaker storm track that does not reach as far north. Atmospheric pressure at sea level is almost always lower at Iceland than Lisbon; it is the magnitude of the difference between them that determines the index.

Controversy about the nature of the pattern was sparked off by the suggestion that the NAO is better described as only part of a larger pattern of variability, which spans the Northern Hemisphere and is centred around the North Pole³. Such a pattern has been identified for the Southern Hemisphere, where it is symmetric about the pole. But in

*The North Atlantic Oscillation, AGU Chapman Conference, Ourense, Spain, 28 November–1 December 2000. Abstracts are at www.ldeo.columbia.edu/NAO/conference/chapman_conf.html

the north the hemispheric mode is dominated by the North Atlantic and Arctic regions, with only a weak link to the north Pacific (C. Deser, Natl Center for Atmospheric Res., Boulder, Colorado)⁴. Which of these views of the NAO is the more useful therefore depends on the research agenda. For instance, for studies of seasonal predictability in the Atlantic region it is usually better to concentrate on the NAO.

Another aspect is that the NAO may not always have shown the same behaviour. Indirect data for the past four centuries suggest that the amplitude of variability might have increased significantly after the end of the Little Ice Age, around 1850 (E. Cook, Lamont-Doherty Earth Obs., Palisades, New York). Moreover, the consequences of the pattern, for example on sea-ice extent in the Arctic (J. E. Walsh, Univ. Illinois, Urbana-Champaign), depend on the exact positions of the two pressure centres — which can differ significantly in years with the same index value.

The persistent positive trend in the index since about 1960 raises the question of whether global warming is responsible. There is no clear answer. A new index, based on surface pressures from Gibraltar, Reykjavik in Iceland and Ponte Delgada in the Azores, shows a similar trend for the early part of the twentieth century (P. Yiou, Lab. des Sciences du Climat et de L'Environ-

nement, Gif-sur-Yvette), implying that natural variability can account for Europe's mild and wet winters since 1960. But an analysis of the Lisbon-Iceland index indicates that natural variability can be rejected as a cause for the recent trend (S. B. Feldstein, Univ. Pennsylvania). After considering these conflicting results and the processes that drive variations of the NAO, most participants, asked to give their best guess for the coming decade, expected the index to increase further, rather than decline towards the mean as would be expected in the absence of a link with climate change.

Two main candidate mechanisms emerged that might explain the long-term persistence of trends in the lower atmosphere (in which disturbances can normally be tracked for only about a week). Interactions with the ocean are one possibility. A statistical analysis of observations reveals two patterns of Atlantic sea surface temperatures that tend to precede specific phases of the NAO: positive anomalies in mid-latitude temperatures induce a positive phase with a lead time of six months, and the tropical Atlantic surface temperatures exert a weaker influence on timescales of up to two months (A. Czaja, Massachusetts Inst. Technol.).

The other candidate mechanism is atmospheric, and centres on the polar stratospheric vortex — a rotating circulation pattern in the higher levels of the atmosphere above the poles. Depending on its strength, the vortex acts as a window that allows atmospheric waves originating in the troposphere to pass into the upper atmosphere, or as a mirror reflecting the waves back down. This second phenomenon could amplify a positive NAO phase if there is interference between the reflected waves and the original waves (H.-F. Graf, MPI für Meteorologie, Hamburg)⁵.

Both sea surface temperatures and the strength of the polar stratospheric vortex are thought to increase with increasing green-

house gases in the atmosphere. So this consideration points to a causal connection between global warming and the positive trend of recent years. If this is indeed the case, there could be a beneficial effect: simulations with a general circulation model suggest that a positive NAO may temporarily mitigate one of the most worrying possible impacts of global warming, the slowing of the thermohaline circulation in the North Atlantic (T. L. Delworth, Geophys. Fluid Dynamics Lab., Princeton Univ.). This circulation is one of the main engines of heat transport around the globe — if it slowed or stopped, the climatic consequences would be severe.

The NAO has a considerable impact on natural ecology and human economies, in arenas ranging from plant and animal population dynamics to energy production in Scandinavia. So many researchers are attempting to predict its behaviour on both seasonal and decadal timescales. At the seasonal scale, the aim is to exploit the predictive value of sea surface temperatures. According to one forecast, in which observed Atlantic sea surface temperatures for May and November 2000 were fed into a model, there is a 66% chance of a positive NAO for the coming winter (M. Rodwell, Hadley Centre, Reading, Berks.). If increasing greenhouse-gas concentrations in the atmosphere are driving the current positive phase, we should expect further increases in the frequency of mild, wet winters in northern Europe. But to confirm whether global warming does indeed change the balance between the positive and negative phases, we will simply have to wait until a clear, tell-tale signal emerges from the noise of natural variability. ■

Heike Langenberg is an associate editor at Nature.

1. Hurrell, J. W. *Science* **269**, 676–679 (1995).
2. Thompson, D. W. J. & Wallace, J. M. *Geophys. Res. Lett.* **25**, 1297–1300 (1998).
3. Walker, G. T. & Bliss, E. W. *Mem. R. Meteorol. Soc.* **4**, 53–84 (1932).
4. Deser, C. *Geophys. Res. Lett.* **27**, 779–782 (2000).
5. Perlwitz, J. & Graf, H.-J. *J. Clim.* **8**, 2281–2295 (1995).

Nanotechnology

Solid progress in ion conduction

Alan V. Chadwick

Materials that conduct ions are useful in devices involving electrochemical reactions, such as fuel cells and batteries. Low ionic conductivity was a problem for these materials until researchers built nanoscale versions.

The development of materials with dimensions in the nanoscale range is a highly active area of research because they often have very different properties from the bulk material. These properties can offer new or improved technological applications. So far, materials scientists interested in nanotechnology have mostly focused on semiconductors and ceramics. For example,

semiconductors made from nanocrystals have unusual optical, electrical and magnetic properties, and ceramics made from nanoparticles have greater hardness and plasticity than normal. However, on page 946 of this issue, Maier and colleagues¹ demonstrate that restructuring simple ionic crystals at the nanoscale can also alter their electrical properties, pointing the way to potentially

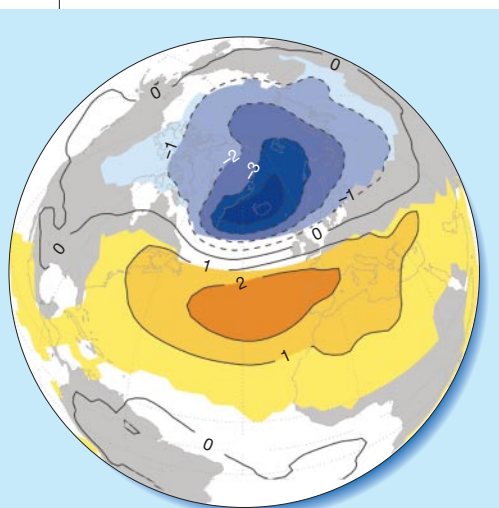


Figure 1 Deviations from a 52-year average of sea-level atmospheric pressures over the North Atlantic region, during winters with a high North Atlantic Oscillation index (1 standard deviation above the average). Anomalously large differences between pressures over Iceland and Lisbon lead to a strong storm track in the North Atlantic, bringing mild, wet and stormy weather to northern Europe. The data are derived from measurements for the months December to March, for 1948–99. Units are hectopascals (= millibars). (Graphic courtesy of Thomas Jung and Michael Hilmer.)

dramatic improvements in energy storage and generation.

In purely ionic solids, such as a crystal of sodium chloride, there are virtually no free electrons, so the migration of charge is restricted to the movement of ions. Such movements are possible because of defects in the crystal lattice — either lattice vacancies (empty lattice sites that are not occupied by ions) or interstitial ions (extra ions between the normal lattice sites). Ions next to a vacancy can jump into it, or an interstitial ion can jump to a neighbouring interstitial site. A series of such jumps allows charge to move through the solid, producing an ionic current.

Lattice defects in a simple ionic crystal only occur in pairs because the overall electrical neutrality of the crystal must be maintained. For example, in NaCl, vacancies on the Na^+ sublattice must be compensated by vacancies on the Cl^- sublattice and vice versa. The concentration of defects in a crystal of NaCl is low because the number of defects is thermodynamically controlled and the energy required to form pairs of them is high. The concentration of defects is only about 0.1% at the melting point, and falls exponentially with decreasing temperature. In addition, the activation energy for an ion, or defect, to jump to a neighbouring site is high, leading to low mobility. So the overall conductivity, which is the product of defect concentration and mobility, is low — in a simple material such as NaCl, the electrical conductivity at the melting point is eight orders of magnitude less than in a metal, and is almost immeasurable at room temperature.

Electrochemists in the nineteenth century recognized that ionic solids could be used as electrolytes in batteries and chemical sensors. But the extremely low conductivity of

most ionic materials makes such devices impractical. As early as 1835, Faraday noted that solid silver sulphide and lead fluoride have unusually high conductivities. A few materials even approach the conductivities of aqueous electrolyte solutions.

In the 1960s this type of material began to attract serious attention with the discovery of new compounds with high ionic conductivities, and the demonstration of feasible devices, particularly solid-state batteries, fuel cells and sensors (Fig. 1). Fuel cells based on these materials — known as fast-ion conductors or solid electrolytes — offer a highly efficient and clean method of large-scale energy production. For example, solid-oxide fuel cells use an electrolyte, such as zirconia, that conducts oxygen ions, with hydrogen as the fuel, air as the oxidant and water as the waste product. The oil crisis of the 1970s and the need for better energy management led to increased activity in this area, and interest has continued to the present day.

The low ionic conductivity, particularly for solids that conduct with lithium ions (for battery electrolytes) and oxygen ions (for fuel cells), still inhibits the commercial development of these devices. One strategy to increase the conductivity of the electrolyte is to add components (known as dopants) that preferentially increase the concentration of one type of defect in the pair. Unfortunately, there is often an optimum level of doping beyond which the conductivity begins to decrease. At the optimal level, the only way to increase conductivity further is to raise the operating temperature to enhance the mobility of the defects. For example, most fuel cells become efficient only at operating temperatures around 1,000 °C. An alternative is to find crystal structures that favour the formation and

Daedalus

David Jones

David Jones, author of the Daedalus column, is indisposed.

migration of defects. This has had some limited success, but the long-term mechanical stability of these materials can be a serious drawback.

The surface properties of an ionic crystal create a 'space-charge region', in which the concentration of defects at the surface is governed by the energies of individual defects and one type of defect in a pair can dominate. As a result, the surface has an excess of one type of defect, as happens with chemical doping, and ionic conduction in this region is accordingly enhanced. This is seen experimentally² when simple ionic crystals are mixed with insulators to produce samples with many interfacial regions. Theoretical predictions (including previous work by Maier³) indicate that there should be dramatic increases in ionic conductivity when the spacing of interfaces is comparable to, or smaller than, the space-charge layer.

In the new experiment, Maier and colleagues¹ create alternating thin layers of two simple ionic crystals, barium fluoride and calcium fluoride. They find that the conductivity along the layers (that is, parallel to the interfaces) shows a marked increase as the thickness of the layers decreases. When the thickness of the calcium fluoride–barium fluoride layer is just 16 nanometres — which is when the space-charge layers begin to overlap at the centre of the layers — the conductivity becomes between 100 and 1,000 times higher than in bulk crystals of barium fluoride.

This work is the first definitive proof of increased ionic conductivity in nanoscale materials. The general approach should also apply to other ionic solids, so it offers a predictable way to design a wide range of new ionic materials with practical applications. An obvious technological target is the preparation of nanoscale lithium-ion conductors to make more efficient rechargeable batteries for portable electronics (Fig. 1). There are still technical problems to be resolved, such as the scaling up of laboratory techniques for mass production. Nonetheless, this result suggests that the emerging field of nano-ionics has a healthy future.

Alan V. Chadwick is at the Centre for Materials Research, School of Physical Sciences, University of Kent, Canterbury, Kent CT2 7NR, UK.

e-mail: a.v.chadwick@ukc.ac.uk

1. Sata, N., Eberman, K., Eberl, K. & Maier, J. *Nature* **408**, 946–949 (2000).

2. Liang, C. C. J. *Electrochem. Soc.* **120**, 1289–1292 (1973).

3. Maier, J. *Prog. Solid State Chem.* **23**, 171–263 (1995).

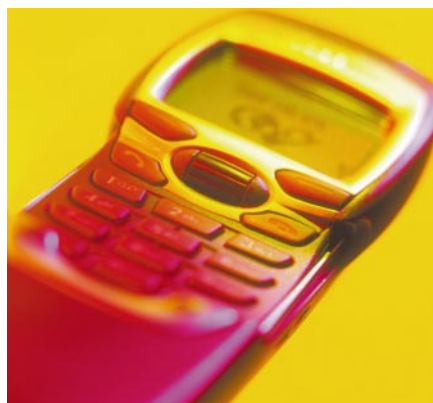
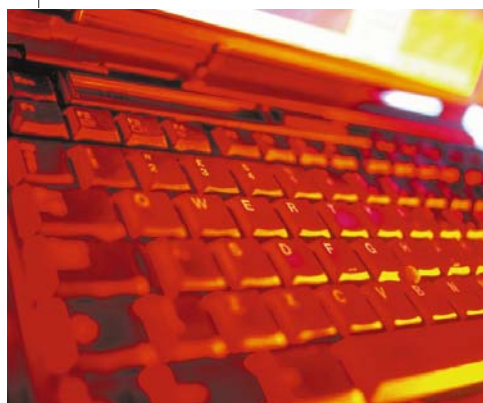


Figure 1 A possible target for research into ionic conductors. Apart from solid-state fuel cells and ion sensors, lithium-ion technology is being developed for use in rechargeable batteries. Lithium-ion batteries have large capacity and high energy output, ideal for applications in electrical vehicles and in energy storage (storing excess electricity produced at power stations during 'off-peak' periods). They also make attractive batteries for portable electronic devices, such as mobile phones and laptop computers, which require long lifetimes. The material created by Maier and colleagues¹ has alternating layers of ionic crystals — just nanometres thick — that improve their conductivity by several orders of magnitude.

Does the Queen speak the Queen's English?

Elizabeth II's traditional pronunciation has been influenced by modern trends.

The pronunciation of all languages changes subtly over time¹, mainly owing to the younger members of the community². What is unknown is whether older members unwittingly adapt their accent towards community changes. Here we analyse vowel sounds from the annual Christmas messages broadcast by HRH Queen Elizabeth II during the period between the 1950s and 1980s. Our analysis reveals that the Queen's pronunciation of some vowels has been influenced by the standard southern-British accent of the 1980s which is more typically associated with speakers who are younger and lower in the social hierarchy.

Phoneticians have documented many types of change to the standard accent of British English known as 'received pronunciation'³, some of which have a corollary in the changing attitudes towards social class. There was a marked social stratification in Britain in the 1950s⁴, and in 1963 the phonetician David Abercrombie wrote, "One either speaks received pronunciation, or one does not, and if the opportunity to learn it in youth has not arisen, it is almost impossible to learn it in later life"⁵. But as class distinctions have become more blurred⁴, so too have the boundaries between English accents that mark social class.

Although modern received pronunciation has resisted many of the stigmatized features of the London cockney accent, such as 'h' dropping, it has nevertheless been influenced by cockney — for example in the tendency to pronounce the 'l' in 'milk' as a vowel⁶. Some of these changes in pronunciation in England have been led by younger members of the population, who reject received pronunciation because of its association with the Establishment⁷ — much to the chagrin of the older generation.

But can the traditional accent of older members of the community be preserved against such influences? And if not, is it still realistic to define supposedly immutable pronunciation standards⁸ towards which the community should strive?

We investigated this issue by acoustic analysis (with the permission of Buckingham Palace) of the vowels from the Christmas messages broadcast every year by the Queen since 1952, comparing the vowel sounds from the 1950s with those from the 1980s. The BBC provided us with the recordings from their archives.

We also analysed whether there had been any change towards a 1980s standard southern-British (SSB) accent, which is similar in many ways to the accent of the

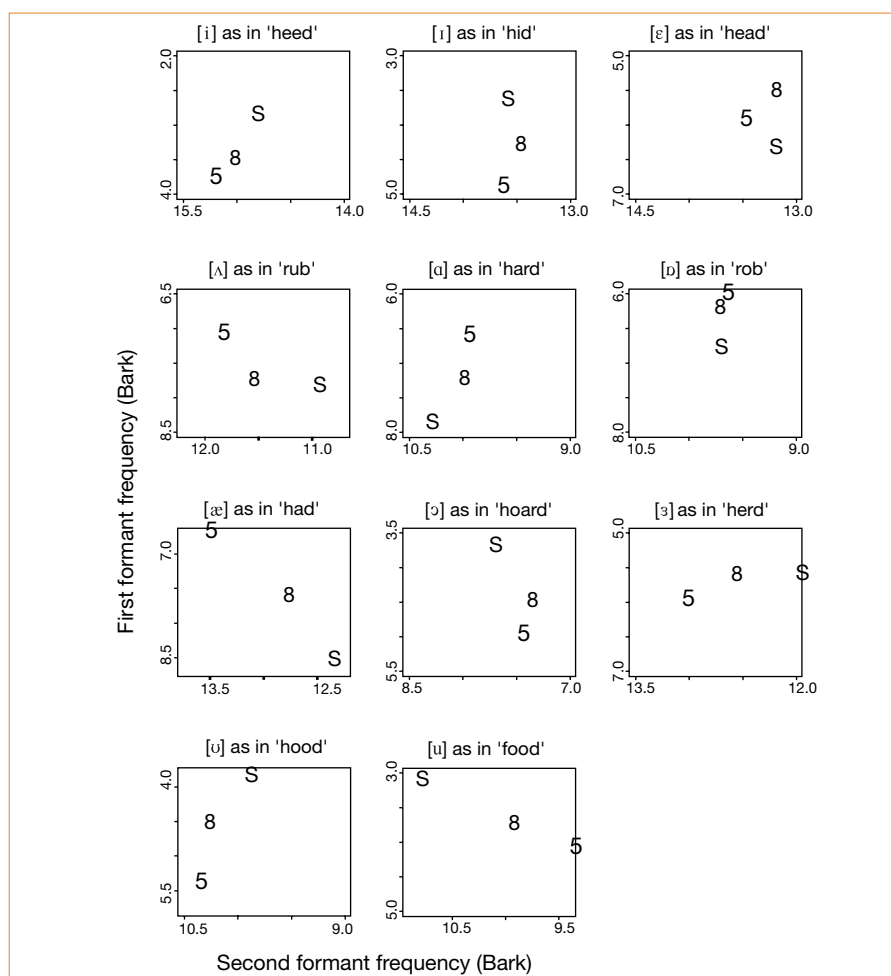


Figure 1 The three symbols '5', '8' and 'S' represent the average positions of different vowel types in the Christmas broadcasts of the 1950s and 1980s, and in standard southern British of the 1980s, respectively. The axes are the first two formant frequencies in Bark, a scale used to model the way listeners perceive vowels¹². Positions towards the top of each square correspond to less mouth opening; the left corresponds to sounds made by constricting the vocal tract nearer the lips rather than further back¹¹.

Queen, but more likely to be spoken by most of the middle classes and by younger speakers. The SSB data⁹ were taken from an existing corpus of female BBC broadcasters recorded in the 1980s¹⁰. Our acoustic analysis of the extent and direction of vowel changes in these three data sets was based on a well-established procedure of calculating the first two resonances or formant frequencies of the vocal tract¹¹.

Our results show that there were significant changes in at least one formant for 10 of 11 vowel sounds and in both formants for 5 of 11 vowel sounds from the 1950s to the 1980s Christmas broadcasts. Moreover, the average position of the 1980s vowels in the formant space is between those of the 1950s and SSB positions (Fig. 1). These results indicate that the vowels in the Christmas message have moved towards, but not attained, their SSB equivalents from

the 1980s. Thus, there has been a drift in the Queen's accent towards one that is characteristic of speakers who are younger and/or lower in the social hierarchy.

We conclude that the Queen no longer speaks the Queen's English of the 1950s, although the vowels of the 1980s Christmas message are still clearly set apart from those of an SSB accent. The extent of such community influences is probably more marked for most adult speakers, who are not in the position of having to defend a particular form of English (the Queen's English in this case). The chances of societies and academies successfully preserving a particular form of pronunciation against the influence of community and social changes are as unlikely as King Canute's attempts to defeat the tides.

Jonathan Harrington*, Sallyanne Palethorpe*, Catherine I. Watson†

*Macquarie Centre for Cognitive Science and
†Speech Hearing and Language Research Centre,
Macquarie University, Sydney,
New South Wales 2109, Australia
e-mail: jmh@shlrc.mq.edu.au

1. Aitchison, J. *Language Change: Progress or Decay?* (Cambridge Univ. Press, 1991).
2. Labov, W. *Principles of Linguistic Change* (Blackwell, Oxford, 1994).
3. Wells, J. *Accents of English* (Cambridge Univ. Press, 1982).
4. Cannadine, D. *Class in Britain* (Yale Univ. Press, New Haven, 1998).

5. Abercrombie, D. *Problems and Principles in Language Study* (Longman, London, 1963).
6. Wells, J. in *Nonstandard Varieties of Language* (eds Melchers, G. & Johannesson, N.-L.) 198–205 (Almqvist & Wiksell, Stockholm, 1994).
7. Gimson, A. & Cruttenden, A. *Gimson's Pronunciation of English* (Arnold, London, 1994).
8. Honey, J. *Language is Power* (Faber, London, 1997).
9. Deterding, D. J. *Int. Phonet. Assoc.* 27, 47–55 (1997).
10. Roach, P. et al. *J. Int. Phonet. Assoc.* 23, 47–54 (1994).
11. Ladefoged, P. *Three Areas of Experimental Phonetics* (Oxford Univ. Press, 1967).
12. Johnson, K. *Acoustic and Auditory Phonetics* (Blackwell, Oxford, 1997).

Ecophysiology

Penguin fathers preserve food for their chicks

The king penguin *Aptenodytes patagonicus* feeds only at sea and must live off its reserves when it comes ashore to breed. We found that male penguins returning to their egg between three weeks before and ten days after it hatches bring food for the chick in their stomachs. This food can be preserved in the stomach for two to three weeks while the male fasts, enabling him to feed the chick if the female's return is delayed.

A penguin pair take it in turns to incubate their egg during the 54 days before it hatches¹. The non-incubating partner goes

off to sea to feed, mainly on myctophid fish². This entails a journey of 400–500 km south from the birds' colonies in the Crozet Archipelago³, which, combined with the variable availability of food^{4,5}, makes the duration of foraging trips unpredictable. Either mate may therefore be with the egg when it hatches⁶. To survive, the newborn chick needs regurgitated food from the parent in attendance, so how does a breeder cope with a delay in its mate's return?

We measured the changes in the stomach contents of breeding king penguins in relation to four hatching schedules (Fig. 1). After laying, the females went to sea with an empty stomach (weight of stomach contents, 24 ± 4.5 g; $n = 5$) while the males took over the incubation. When females returned more than one month before

hatching, their stomachs were empty apart from a few pebbles (41 ± 9.8 g; $n = 10$). Relieved males left with empty stomachs for the sea (22 ± 3.6 g; $n = 5$); the length of this foraging trip varied considerably. Returning males had more in their stomachs the closer they arrived to the date of hatching ($r = 0.64$, $P < 0.001$, $n = 31$) but less if they came back later ($r = -0.57$, $P = 0.003$, $n = 28$) (Fig. 2).

Ninety per cent of males that returned more than 10 days before hatching (Fig. 1a) had food in their stomachs (stomach contents, 210 ± 28.0 g; $n = 18$). By hatching time, 40% of all males had not been relieved by the returning female and so needed to give this food to the chick (Fig. 1b, c). During three weeks of incubation, males lost 160 g in weight per day, amounting to 20% of their initial body mass, but the amount of food in the stomachs of relieved males (248 ± 32.3 g; $n = 21$) was not significantly different from that of birds arriving at the colony (Mann–Whitney test, $U = 115.5$; $P > 0.05$).

Prey that had been stored in the stomach for 20 days was in a similar state of preservation to that in penguins arriving from the sea — fish remains were mashed up and squid parts intact. But the proportions of lipid and protein in the dry stomach contents differed (lipid: arrivers, $23 \pm 1.2\%$; leavers, $16 \pm 1.1\%$; Mann–Whitney test, $U = 44$; $P < 0.01$; protein: arrivers, $56 \pm 1.7\%$; leavers, $68 \pm 0.8\%$; $U = 14$; $P < 0.01$; 17 arrivers and 13 leavers).

Based on the energy requirements of a newly hatched chick⁷, we found that there was enough energy ($1,660 \pm 156$ kJ, $n = 30$) in the stomach contents to sustain a newly hatched chick for about ten days. This fits with our observation that five of the 12 males that left with an empty stomach fed their chick for $9 (\pm 1.3)$ days.

Breeding king penguins evidently have some kind of internal clock that tells them to return from the sea bearing food only if their arrival falls within the hatching schedule and fits with the probability that the mate has not yet deserted. This storage and conservation of food in the stomach during several weeks of fasting, in anticipation of a possible delay

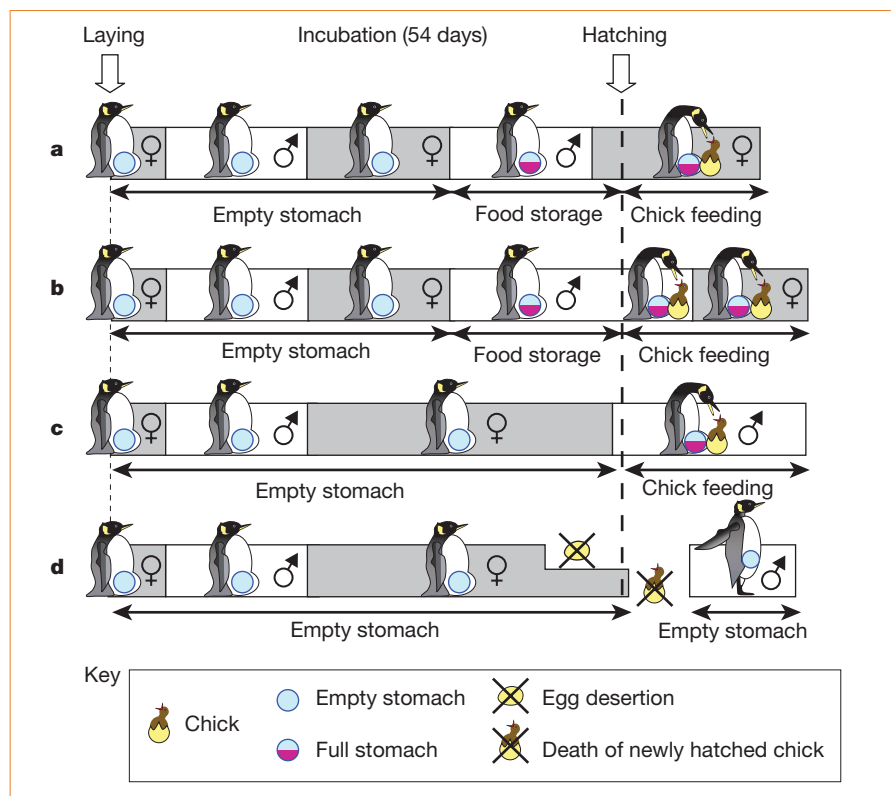


Figure 1 Stomach content of king penguins from Possession Island, Crozet Archipelago, in relation to the date of egg hatching and possible outcome of incubation due to the variable duration of the mate's time at sea. **a**, The female usually relieves the male a few days before hatching and his stored food is not needed; **b**, the chick hatches before the female returns and the male can feed it with the food stored in his stomach; **c**, arriving just before or soon after hatching, the male brings more food in his stomach (Fig. 2); **d**, the female has deserted the egg or the chick has died from starvation and the male, arriving long after the hatching date, has an empty stomach.

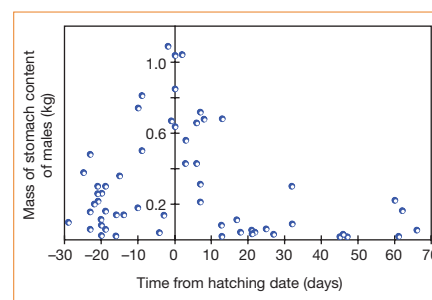


Figure 2 Relation between the mass of the stomach contents of male king penguins returning to the colony for the fourth shift and the time of egg hatching. Stomachs were flushed and their contents drained and weighed⁸. The bird was fed after stomach sampling so as not to disrupt chick feeding.

in the female's return, is presumably an adaptation to the large but highly variable marine resources of the austral ocean.

Michel Gauthier-Clerc, Yvon Le Maho, Yannick Clerquin, Samuel Draut, Yves Handrich

Centre d'Écologie et Physiologie Énergétiques, Centre National de la Recherche Scientifique, 23 rue Becquerel, 67087 Strasbourg cedex 2, France, and Institut Français pour la Recherche et la Technologie Polaire, 29280 Plouzané, France
e-mail: yvon.lemaho@c-strasbourg.fr

1. Weimerskirch, H., Stahl, J. C. & Jouventin, P. *Ibis* **134**, 107–117 (1992).
2. Cherel, Y. & Ridoux, V. *Ibis* **134**, 118–127 (1992).
3. Jouventin, P. et al. *Mar. Ecol. Prog. Ser.* **106**, 11–19 (1994).
4. Foxton, P. *Disc. Rep.* **28**, 191–236 (1956).
5. Hart, T. J. *Disc. Rep.* **21**, 261–356 (1942).
6. Barrat, A. *Com. Nat. Fr. Rech. Ant.* **40**, 9–52 (1976).
7. Barré, H. *J. Physiol. (Paris)* **74**, 555–561 (1978).
8. Wilson, R. P. *J. Field Ornithol.* **55**, 109–112 (1984).

Biomechanics

Penguin waddling is not wasteful

Penguins use twice as much metabolic energy as other terrestrial animals of a similar mass to walk a given distance^{1,2}, which was thought to be because side-to-side waddling requires excessive work. Here we show that waddling actually conserves mechanical energy and suggest instead that walking is expensive for penguins because their short legs require them to generate muscular force rapidly.

Many animals reduce the muscular work of walking by exchanging the gravitational potential energy and kinetic energy of the centre of mass, like an inverted pendulum³. But penguins' waddling gait appears to involve large fluctuations in lateral kinetic energy and a poor exchange of mechanical energy¹. To test the idea that penguin walking is expensive because it requires more mechanical work to lift and accelerate their body than other animals of a similar mass, we used a force platform to measure the vertical, fore–aft and lateral ground-reaction forces of five walking emperor penguins (*Aptenodytes forsteri*). Forces were integrated to calculate the centre-of-mass velocities, displacements, and fluctuations in gravitational potential energy and kinetic energy⁴.

The fluctuations in lateral velocity were nearly three times those in the fore–aft direction. Lateral-velocity fluctuations occurred about zero, whereas the fore–aft velocity fluctuated about the average walking speed. Thus, kinetic energy fluctuates much less laterally than in the fore–aft direction. At moderate walking speeds of about 0.5 m s^{−1}, fluctuations in lateral kinetic energy accounted for about 30% of the total kinetic-energy fluctuations — a large value compared with other terrestrial vertebrates^{5,6}.

Penguins recovered up to 80% of the

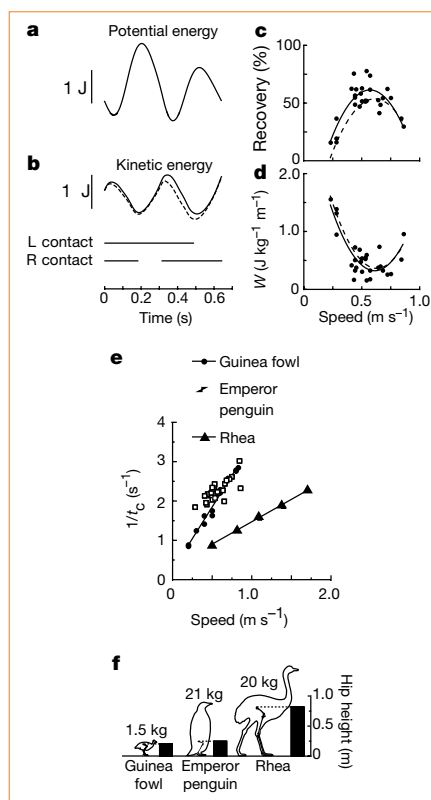


Figure 1 Biomechanics of penguin walking. **a,b**, Fluctuations in gravitational potential and kinetic energy of the centre of mass for an emperor penguin walking at 0.54 m s^{−1}. Bars, times when left or right foot was in contact with the ground. J, joules. **c,d**, Inverted pendulum recovery of mechanical energy (**c**) and work (**d**) per unit distance at a range of walking speeds. **e**, Rate of generating muscular force, as indicated by the inverse of ground-contact time (1/t_c), for emperor penguin, guinea fowl (*Numida meleagris*)¹⁰ and rhea (*Rhea americana*) (P. G. Weyand, unpublished results), and **f**, standing hip height¹¹ (leg length). The metabolic costs of transport of emperor penguins (8.64 J kg^{−1} m^{−1}) and guinea fowl (8.38 J kg^{−1} m^{−1}) are twice those of rheas (4.23 J kg^{−1} m^{−1}) (ref. 8).

mechanical energy during each stride, because of the similar amplitudes and out-of-phase fluctuations in gravitational potential and kinetic energies (Fig. 1a–c). This value is among the highest reported for any animal, indicating a large reduction in the mechanical work done by the penguin's muscular system (Fig. 1d). Recovery of mechanical energy was greatest and the work performed on the centre of mass per distance walked was least at about 0.5 m s^{−1}, a speed commonly used in nature².

Surprisingly, excluding lateral kinetic energy — that is, waddling — from our calculations would result in the recovery of less mechanical energy and more work being required from the muscles (Fig. 1b–d). Lateral movements increase the kinetic energy available to convert into gravitational potential energy and also cause these energies to fluctuate more completely out of phase.

Other components of work, such as rotating the body about the centre of mass

and swinging the limbs, probably do not account for the high cost: rotational work is approximately 10% of the total centre-of-mass work, and penguins have comparatively light, short limbs which offset their quick strides. At their preferred speed, emperor penguins perform the same amount of mechanical work to lift and accelerate their centre of mass as other walking animals of the same mass. Waddling therefore does not explain the high metabolic cost of penguin walking.

We propose instead that this high cost is due to rapid force generation. Regardless of the work muscles perform, they must actively generate force to support body-weight, and this cost is proportional to the magnitude and rate of generating force^{7,8}. At the same walking speed, emperor penguins and rheas — a flightless bird of a similar mass — generate the same average force on the ground. But because of their comparatively short legs, emperors generate this force twice as fast (Fig. 1e, f), requiring faster, less economical muscle fibres⁹. Our results indicate that it is their short legs, and not their waddling gait, that accounts for the high cost of penguin walking.

Timothy M. Griffin*, Rodger Kram†

*Integrative Biology Department, University of California, Berkeley, California 94720-3140, USA

†Department of Kinesiology and Applied Physiology, Campus Box 354, University of Colorado, Boulder, Colorado 80309-0354, USA
e-mail: tmgriffin@uclink4.berkeley.edu

1. Pinshow, B., Fedak, M. A. & Schmidt-Nielsen, K. *Science* **195**, 592–594 (1977).
2. Dewasmes, G., Le Maho, Y., Cornet, A. & Groscolas, R. *J. Appl. Physiol.* **49**, 888–896 (1980).
3. Cavagna, G. A., Heglund, N. C. & Taylor, C. R. *Am. J. Physiol.* **233**, R243–R261 (1977).
4. Cavagna, G. A. *J. Appl. Physiol.* **39**, 174–179 (1975).
5. Farley, C. T. & Ko, T. C. *J. Exp. Biol.* **200**, 2177–2188 (1997).
6. Cavagna, G. A., Saibene, F. P. & Margaria, R. *J. Appl. Physiol.* **18**, 1–9 (1963).
7. Kram, R. & Taylor, C. R. *Nature* **346**, 265–267 (1990).
8. Roberts, T. J., Kram, R., Weyand, P. G. & Taylor, C. R. *J. Exp. Biol.* **201**, 2745–2751 (1998).
9. Rall, J. A. *Exercise Sport Sci. Rev.* **13**, 33–74 (1985).
10. Gatesy, S. M. *J. Morphol.* **240**, 115–125 (1999).
11. Gatesy, S. M. & Biewener, A. A. *J. Zool.* **224**, 127–147 (1991).

Ecology

Mistletoe seed dispersal by a marsupial

The temperate forest that extends from 35° S to 55° S along the Pacific rim of southern South America is home to an endemic and threatened flora and fauna¹. Many species belong to lineages that can be traced back to ancient Gondwanaland^{2,3}, and there are some unusual interactions between plants and animals. Here we describe an exclusive association that involves the dispersal of the sticky seeds of a mistletoe by a marsupial, *Dromiciops australis*, endemic to this region — a task

previously thought to be carried out exclusively by birds.

Birds have been described as the animal dispersers of the shrubby stem-parasitic plants of the two most species-rich mistletoe families, Loranthaceae and Viscaceae^{4,5}. The mistletoe's aerial parasitic lifestyle imposes stringent requirements on reproduction by seed, so it is not surprising that bird–mistletoe interactions are highly specialized⁶. Efficient dispersal of mistletoes involves not only the ingestion and transport of the sticky seeds produced by these plants, but also their placement on the living branches of an appropriate host.

In the Lake District of southern Argentina, we found that *D. australis* — a nocturnal marsupial endemic to the northern portion of the temperate forests of southern South America⁷ — is the exclusive disperser of the seeds of the loranthaceous mistletoe *Tristerix corymbosus*. Mature fruits of *T. corymbosus* are green (Fig. 1a), a colour often associated with mammalian dispersers⁸.

During 500 hours of observation, we saw no birds eat these fruits. Although we found seeds of different flesh-fruit species in the stomach contents of 61% of 296 mist-netted birds, these did not include any from *T. corymbosus*. We timed the removal of 92 fruits: 91 were removed at night, and only one during the day. Seeds were often deposited on vertical trunks of shrubs and small trees in patterns incompatible with typical bird behaviour (Fig. 1b).

Automatic cameras photographed two species of mammal, both nocturnal, in trees supporting mistletoe plants. Judging by these photographs, *D. australis* was more abundant ($n=101$ photographs) than the rodent *Irenomys tarsalis* ($n=23$). Of these two, apparently only *D. australis* eats and defecates intact seeds of *T. corymbosus* (Fig. 1c). In the field, 84% of 57 live-trapped marsupials defecated 1–31 *T. corymbosus* seeds each (mean, 6.3 seeds) while still in the traps. None of 13 live-trapped *I. tarsalis* defecated mistletoe seeds or consumed fruits while in captivity.

We found that *D. australis* is a highly efficient disperser of mistletoe seeds. Ten captive *D. australis* ate all ten *T. corymbosus* fruits offered to each of them within an hour, peeling off the exocarp with their forepaws and swallowing the rest whole, then defecating 98% of the ingested seeds ($n=653$) undamaged. In the field, we counted 192 seeds in or above seed traps and estimated that only about 10% of the seeds defecated by *D. australis* fall to the ground.

Passage of seeds through the marsupial's gut is critical for germination and development of a 'holdfast' (the disc-like swelling at the end of the radicle that effects its attachment to the host plant). Most seeds removed from the exocarp by hand failed to germinate, neither could they infect host plants.

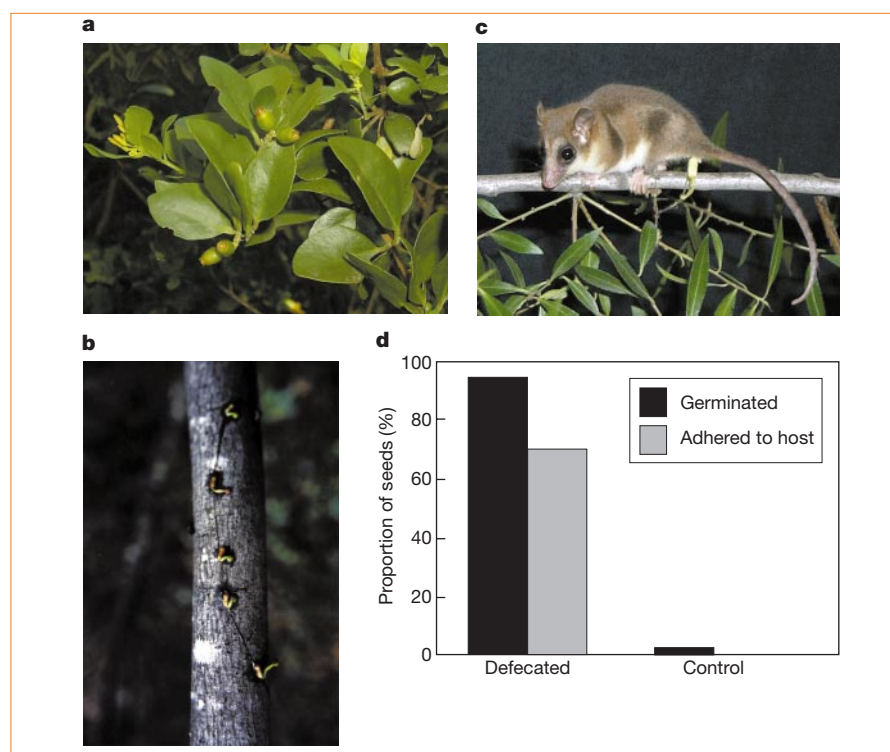


Figure 1 Seed dispersal of *Tristerix corymbosus*. **a**, Mature fruits, which measure about 1×0.6 cm, contain one seed measuring 0.7×0.4 cm. **b**, Germination of naturally dispersed seeds. Several seeds linked by a viscin thread are usually defecated in strings. These strings may stick to vertical trunks and branches, as shown, in spiral patterns around branches, or even to the underside of branches. Unlike other mistletoes which can only establish on thin branchlets^{5,12}, *T. corymbosus* can establish on branches and trunks up to 8 cm wide. **c**, A male *Dromiciops australis* (also known as the monito del monte), about 20 cm long from nose to tail tip, defecates a mistletoe seed. **d**, Percentage of seeds defecated by *D. australis* ($n=200$) and of seeds with the exocarp removed by hand ($n=200$; controls) that germinated and developed a holdfast that adhered to the host. We conducted the study in the Llaolao Reserve, Argentina ($41^{\circ} 8' S$, $71^{\circ} 19' W$), during the 1998–99 and 1999–2000 fruiting seasons (December–March). The site supports an old-growth forest of *Nothofagus dombeyi*. The shrubs *Aristotelia chilensis* and *Maytenus boaria* are the main hosts of *T. corymbosus* in the study area. On 15 sampling dates, we tagged fruits for 24 h and checked them at dawn and at dusk. We mist-netted birds for 1,579 net-hours, and used water to flush the stomachs of all passerines. We live-trapped small mammals using single-door cage traps for 360 trap-nights. We monitored arboreal wildlife using two cameras for 90 d. During 1999–2000, we used six seed traps (2×2 m plastic sheet) to estimate the relative frequencies of naturally dispersed *T. corymbosus* seeds that stuck to plants or fell to the ground. At weekly intervals we counted new seeds falling on the sheets and those sticking to the branches overhead. Full details are available from the authors.

But over 90% of the seeds collected from marsupial faeces germinated. Most of these seeds also developed a holdfast that adhered to the host's bark (Fig. 1d).

Dromiciops australis is the only living representative of the family Microbiotheriidae, a marsupial lineage of presumed Gondwanan origin⁹. The Loranthaceae date back to the early to mid-Cretaceous⁵ and may likewise be of Gondwanan origin. In particular, *Tristerix* is considered to be one of the most primitive genera in this diverse family of nearly 1,000 extant species⁵, so the ancestors of *D. australis* could have been dispersing mistletoe seeds for more than 70 million years, before the breakup of western Gondwanaland. The bird lineages that disperse loranthaceous seeds originated no more than 20–25 Myr ago^{6,10}.

Our findings suggest that marsupial dispersal of mistletoe seeds might represent a very primitive mutualism in the Loranthaceae. Fossils of Palaeocene microbiotheriids found at different localities along the tropical Andes¹¹ may also indicate that the

marsupial–mistletoe mutualism described here was more widespread in the past.

Guillermo Amico, Marcelo A. Aizen

Laboratorio Ecotono, CRUB, Universidad Nacional del Comahue, Unidad Postal Universidad, 8400 San Carlos de Bariloche, Río Negro, Argentina
e-mail: marcito@crub.uncoma.edu.ar

- Armesto, J., Rozzi, R., Smith-Ramírez, C. & Arroyo, M. T. K. *Science* **282**, 1271–1272 (1998).
- Arroyo, M. T. K., Riveros, M., Peñaloza, A., Cavieles, L. & Faggi, A. M. in *High Latitude Rainforests and Associated Ecosystems of the West Coast of the Americas* (eds Lawford, R. G., Alaback, P. B. & Fuentes, E.) 134–172 (Springer, New York, 1996).
- Axelrod, D. I., Arroyo, M. T. K. & Raven, P. *Rev. Chil. Hist. Nat.* **64**, 413–446 (1991).
- Reid, N., Smith, M. S. & Yan, Z. in *Forest Canopies* (eds Lowman, M. D. & Nadkarni, N. M.) 285–310 (Academic, New York, 1995).
- Calder, M. & Bernhardt, P. (eds) *The Biology of Mistletoes* (Academic, New York, 1983).
- Reid, N. *Aust. J. Ecol.* **16**, 457–469 (1991).
- Marshall, L. G. *Mammal. Species* **99**, 1–5 (1978).
- van der Pijl, L. *Principles of Dispersal in Higher Plants* (Springer, Berlin, 1982).
- Springer, M. S. *et al. Proc. R. Soc. Lond. B* **265**, 2381–2386 (1998).
- Tambussi, C. & Noriega, J. *Münchener Geowiss. Abh.* **30**, 245–264 (1996).
- Goin, F. in *Vertebrate Paleontology in the Neotropics* (eds Kay, R. *et al.*) 185–204 (Smithsonian Institution, Washington DC, 1997).
- Sargent, S. *Funct. Ecol.* **9**, 197–204 (1995).

The cosmic microwave background radiation temperature at a redshift of 2.34

R. Srianand*, P. Petitjean†‡ & C. Ledoux§

* IUCAA, Post Bag 4, Ganeshkhind, Pune 411 007, India

† Institut d'Astrophysique de Paris–CNRS, 98bis Boulevard Arago, F-75014 Paris, France

‡ CNRS 173–DAEC, Observatoire de Paris–Meudon, F-92195 Meudon Cedex, France

§ European Southern Observatory, Karl Schwarzschild Strasse 2, D-85748 Garching bei München, Germany

The existence of the cosmic microwave background radiation is a fundamental prediction of hot Big Bang cosmology, and its temperature should increase with increasing redshift. At the present time (redshift $z = 0$), the temperature has been determined with high precision to be $T_{\text{CMBR}}(0) = 2.726 \pm 0.010$ K. In principle, the background temperature can be determined using measurements of the relative populations of atomic fine-structure levels, which are excited by the background radiation. But all previous measurements have achieved only upper limits, thus still formally permitting the radiation temperature to be constant with increasing redshift. Here we report the detection of absorption lines from the first and second fine-structure levels of neutral carbon atoms in an isolated cloud of gas at $z = 2.3371$. We also detected absorption due to several rotational transitions of molecular hydrogen, and fine-structure lines of singly ionized carbon. These constraints enable us to determine that the background radiation was indeed warmer in the past: we find that $T_{\text{CMBR}}(z = 2.3371)$ is between 6.0 and 14 K. This is in accord with the temperature of 9.1 K predicted by hot Big Bang cosmology.

One of the firm predictions of the standard Big Bang model is the existence of relic radiation from the hot phase the Universe experienced at early times¹. The cosmic microwave background radiation (CMBR) was discovered serendipitously² in 1964. The fact that its spectrum follows, with great precision, a planckian distribution over several decades in frequency is a strong argument in favour of the hot Big Bang cosmology. However, the presence of the radiation at earlier times has never been proved directly. The temperature of its black-body spectrum is predicted to increase linearly with redshift $T_{\text{CMBR}}(z) = T_{\text{CMBR}}(0) \times (1 + z)$ and its local value has been determined very accurately by the Cosmic Background Explorer (COBE) to be $T_{\text{CMBR}}(0) = 2.726 \pm 0.010$ K (ref. 3). Detecting the presence of relic radiation at earlier epochs and confirming the well defined predicted temperature evolution is therefore a crucial test for cosmology. To perform this test, we can use the possibility that excited fine-structure levels of the ground-state of atomic species may be partly populated by the cosmic microwave background radiation when the energy separation of the levels is similar to the energy at the peak of the radiation energy distribution.

The relative populations of excited levels can be measured from the absorption lines seen in the spectrum of distant quasars and used to constrain T_{CMBR} at high redshift^{4–10}. The expected values make neutral carbon, C^0 , particularly suitable for this purpose. The ground term is split into three levels ($J = 0, 1, 2$) with the $J = 0 \rightarrow 1$ and $J = 1 \rightarrow 2$ energy separations kT corresponding to, respectively, $T = 23.6$ and 38.9 K. However, the fine-structure levels can also be excited by collisions (mostly with electrons and hydrogen) and by ultraviolet pumping and following cascades. The ionization potential of neutral carbon, 11 eV, is below the ionization potential of hydrogen, 13.6 eV, and C^0 can only be seen in dense, neutral and highly shielded gas⁹. Therefore, excitation by collisions cannot be neglected.

To constrain T_{CMBR} , the kinetic temperature, the particle density of the gas and the local ultraviolet radiation field must be known. As it is very difficult to disentangle these various excitation processes,

all measurements until now have led only to upper limits on T_{CMBR} .

Our measurement was obtained in a unique absorption system where absorption lines of neutral carbon in the three fine-structure levels of the ground term, C^0 , C^{0*} and C^{0**} are observed together with absorption lines of singly ionized carbon in its excited fine-structure level, C^{+*} , and absorption lines of molecular hydrogen (H_2) in the $J = 0$ to 5 rotational levels. The population and depopulation of the first excited rotational level of H_2 ($J = 1$) from and to the ground state ($J = 0$) is controlled by thermal collisions. Therefore the excitation temperature T_{01} is approximately equal to the kinetic temperature. The fine-structure upper level of the C^+ ground-state doublet is mostly populated by collisions and depopulated by radiative decay. Therefore, once the temperature is known, the particle density can be derived from the C^{+*}/C^+ ratio. Finally, the ultraviolet radiation flux can be constrained from the populations of the $J = 4$ and 5 H_2 rotational levels.

Observations

We used the Ultra-violet and Visible Echelle Spectrograph (UVES)¹¹ mounted on the European Southern Observatory (ESO) KUEYEN 8.2-m telescope at the Paranal observatory on 5 and 7 April 2000, to obtain a high-spectral-resolution spectrum of the quasar PKS1232+0815 with emission redshift $z_{\text{em}} = 2.57$ and $m_V = 18.4$. Standard settings have been used in both arms of the spectrograph. Wavelength ranges were 3,290–4,519 Å in the blue and 4,623–5,594 and 5,670–6,652 Å for the red ships. The slit width was 1 arcsec and the charge-coupled devices (CCDs) were binned 2×2 resulting in a wavelength resolution, $\Delta\lambda/\lambda$, of $\sim 45,000$. The exposure time was 3 h in seeing conditions better than 0.8 arcsec full-width at half-maximum. The data were reduced using the UVES pipeline in an interactive mode. The pipeline is a set of procedures implemented in the dedicated context of MIDAS, the ESO data reduction package. Its main tasks are to perform a precise inter-order background subtraction for science frames and master flat-fields, and to allow for an optimal extraction of the object signal rejecting cosmic ray

impacts and performing sky-subtraction at the same time. The reduction is checked step by step. Wavelengths are corrected to vacuum-heliocentric values and individual one-dimensional spectra are combined. This resulted in a signal-to-noise (S/N) ratio per pixel of 10 around 3,700 Å and 20 around 6,000 Å. Typical errors in the wavelength calibration are about 0.5 km s⁻¹.

Fit to the lines, metallicity and molecular content

To our knowledge, this is the first time C⁰, C^{0*} and C^{0**} absorption lines have been detected at high redshift (see Fig. 1). The detection of these three species at an absorption redshift of $z_{\text{abs}} = 2.33771$ is confirmed by the presence of several transitions. C⁰ absorption lines with rest wavelengths 1139.79, 1157.91, 1260.73, 1560.31 and 1656.92 Å, C^{0*} absorption lines with rest wavelengths 1194.40, 1279.89, 1329.09, 1329.10, 1329.12, 1560.68, 1560.71, 1656.27, 1657.38 Å and 1657.91 and C^{0**} lines at 1657.00 and 1658.12 Å are clearly detected and are free from any blending with absorption due to other systems. We use the oscillator strengths given by ref. 12 and standard Voigt profile fitting to determine the amount of matter lying on the line of sight and characterized by column densities, N , and the Doppler parameters, $b = \sqrt{b_{\text{th}}^2 + b_{\text{turb}}^2}$. $b = \sqrt{2k_{\text{B}}T/m}$ is the thermal broadening due to the temperature of the gas, with T the kinetic temperature, k_{B} Boltzmann's constant and m the mass of the particle; b_{turb} is the turbulent broadening due to macroscopic motions of the gas. Figures 1 and 2 show that the system is dominated by a single, well defined component. We consistently impose the same b value on this component for all species and obtain a best-fit for $b = 1.7 \pm 0.1$ km s⁻¹. This value is small compared to the spectral resolution of the data but is ascertained by the relative optical depths in the numerous lines with very different oscillator strengths. Results are given in Table 1 and the fit to a portion of the spectrum is shown in Fig. 1b and in Fig. 2.

The neutral hydrogen column density in the system,

$\log N(\text{H}^0) = 20.90 \pm 0.10$, has been obtained by fitting the damped Lyman α line. At such a high column density, hydrogen is mostly neutral and the elements of interest here, iron, magnesium, silicon and carbon, are mainly in the singly ionized state. Column densities are derived from simultaneous Voigt profile fitting of all the available absorption lines using the same Doppler parameter for all species and the same column density for each species. We concentrate however on weak lines, such as Fe⁺ $\lambda = 1,125$, $\lambda = 1,608$ and $\lambda = 1,611$, Si⁺ $\lambda = 1,808$ or Mg⁺ $\lambda = 1,239$, because they are optically thin. Indeed, in this case, the strength of the line does not depend on the Doppler parameter and gives the column density directly. Profiles of a few transitions are shown on Fig. 2 and results are given in Table 1. The α -chain elements magnesium and silicon have abundances, $Z(X) = N(X)/N(\text{H})$, where X is Mg or Si, that are similar to solar abundances within the measurement uncertainties, $\log Z/Z_{\odot} \approx -1.2$, where $Z_{\odot}$ is the solar abundance. However, we note that, relative to the Sun, iron is underabundant by about a factor of 5 compared to silicon and nickel is slightly underabundant compared to iron. If these differences are because iron and nickel are depleted into dust grains, the depletion is of the same order as that observed in warm halo gas in our Galaxy¹³. All this suggests that the system is very much like typical damped Lyman α systems, with metallicity about one tenth of solar and small depletion of iron type element into dust^{8,14}.

Our echelle data confirms the presence of molecular hydrogen, first detected in ref. 15. At the resolution of the spectrum, most of the H₂ lines are free from blending. The wide wavelength coverage of the spectrum implies that absorption lines from various Lyman bands are seen. Column densities of H₂ in different rotational J levels are obtained by simultaneously fitting various Lyman bands which are free from contamination owing to intervening Lyman α absorption from the diffuse intergalactic medium. Absorption

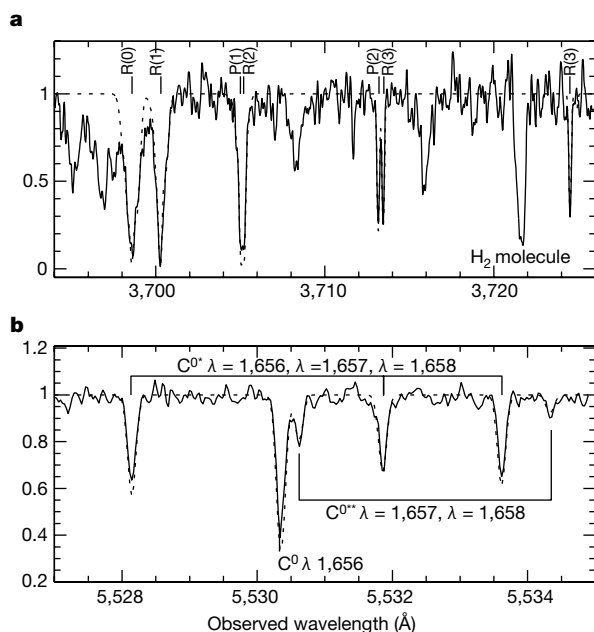


Figure 1 A sample of H₂ and C⁰ absorption lines at $z_{\text{abs}} = 2.33771$. Portions of the normalized spectrum of the quasar PKS1232+0815 taken with the Ultra-violet and Visible Echelle Spectrograph mounted on the 8.2-m KUEYEN telescope of the European Southern Observatory on the Paranal mountain in Chile. **a**, A selection of H₂ absorption lines from the $J = 0, 1, 2$ and 3 rotational levels from the $\nu = 0-1$ Lyman band. The model fitted with the parameters given in Table 2 is overplotted to the data as a dashed line. **b**, Detection of absorption lines from C⁰, C^{0*} and C^{0**} at $z_{\text{abs}} = 2.33771$ in the damped Lyman α system. The model fitted with the parameters given in Table 1 is overplotted to the data as a dashed line.

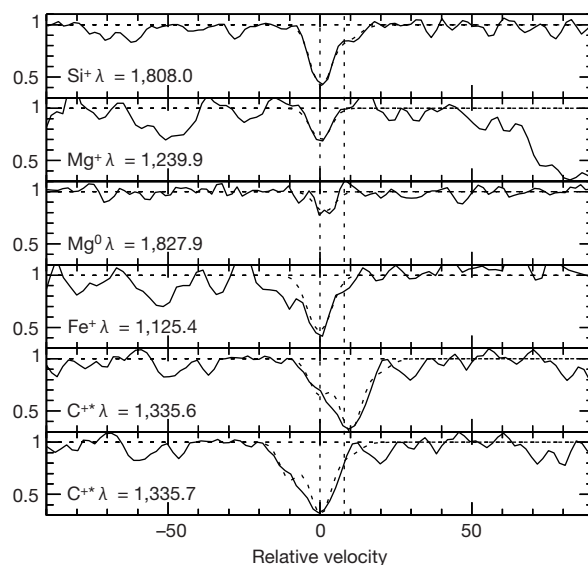


Figure 2 A sample of heavy-element absorption lines at $z_{\text{abs}} = 2.33771$. Absorption profiles of a few transitions (indicated on each panel) from the damped Lyman α system toward PKS1232+0815. The normalized flux is given in a velocity scale with origin at $z = 2.33771$. The C⁺* profile is slightly broader than the other lines. The C⁺* line is a blend of C⁺* $\lambda = 1,335.71$ and C⁺* $\lambda = 1,335.66$. Fitted models are overplotted as dashed lines. The fit is performed over all the absorption lines available in the spectrum. Our best-fitted column density of 10^{14} cm⁻² for C⁺* overpredicts the absorption at $\lambda = 1,335.6$.

profiles corresponding to transitions from $J > 1$ rotational levels are consistent with a single component at the same redshift $z_{\text{abs}} = 2.33771$ as the C^0 lines. Lines from the $J = 0$ and 1 levels are broader and most of the absorption features have similar profiles, suggesting the presence of several components. We obtain a best fit to these absorption lines with four components. Results of Voigt profile fitting are given in Table 2. The total H_2 column density is $N(\text{H}_2) = 1.52 \times 10^{17} \text{ cm}^{-2}$ and the molecular fraction is $f = 2N(\text{H}_2)/(2N(\text{H}_2) + N(\text{H}^0)) = 3.8 \times 10^{-4}$ which is about two orders of magnitude less than what was derived from low dispersion data.

Physical conditions in the gas

In molecular clouds, the kinetic temperature of the gas is given by the excitation temperature, T_{01} , measured between the $J = 0$ and 1 levels. For the components at $z_{\text{abs}} = 2.33735$, 2.33754 and 2.33793 we find $T_{01} = 74 \pm 7$, 64 ± 7 K and 66 ± 10 K respectively. None of these components shows detectable absorption due to C^0 . The characteristics of the gas are very similar to those measured recently in diffuse gas in the halo of our Galaxy¹⁶. For the $z_{\text{abs}} = 2.3371$ cloud, $T = T_{01} = 185 \pm 100$ K. The excitation temperature T_{ex} for other levels, $J = 2$ to 5, is close to 400 K. This means that processes other than collisions are at play in determining the relative populations of the different levels. Indeed, the $J = 4$ and 5 levels can be populated by cascades following ultraviolet pumping and H_2 formation. We can evaluate the ultraviolet pumping rate from $J = 0$ to $J = 4$, $\beta(0)$, using a simple model¹⁷ and assuming that molecules are formed on dust grains with formation rate R .

$$p_{4,0}\beta(0)n(\text{H}_2, J=0) + 0.19Rn(\text{H})n = A(4 \rightarrow 2)n(\text{H}_2, J=4) \quad (1)$$

where $n = n(\text{H}) + 2n(\text{H}_2)$, $p_{4,0} = 0.26$ is the pumping efficiency from level $J = 0$ to level $J = 4$. We use the spontaneous transition probabilities $A(4 \rightarrow 2) = 2.8 \times 10^{-9} \text{ s}^{-1}$ as given by ref. 15. Scaling the formation rate of H_2 in our Galaxy ($R \approx 3 \times 10^{-17} \text{ s}^{-1} \text{ cm}^3$)

with the dust-to-hydrogen ratio measured in the system ($\leq 10^{-1.3}$ the Galactic one) it is easily seen that populating the $J = 4$ level after the formation of a molecule on dust grain is negligible for densities less than $5 \times 10^3 \text{ cm}^{-3}$. We estimate the density from the excitation of C^+ to be at least a factor of 100 less than this value (see below). Moreover, for such large densities, the $J = 0$ and $J = 2$ levels as well as the $J = 1$ and $J = 3$ levels should be in thermal equilibrium, which is not the case¹⁹. We can therefore neglect the population of $J = 4$ and 5 levels by cascades after the formation of a molecule. Finally, we derive from equation (1) that the photo-absorption rate in the Lyman and Werner bands on the surface of the cloud is of the order of $2 \times 10^{-10} \text{ s}^{-1}$, which is quite modest and similar to that observed along sight-lines in our Galaxy¹⁷.

The excited level of the C^+ ground-state term is populated by collisions with electrons and hydrogen atoms²⁰. The electrons cannot be neglected because the corresponding collisional cross-section is large. To estimate the electron density we use the ratio $N(\text{Mg}^+)/N(\text{Mg}^0) = 178$ which is fairly well determined from the detection of the weak lines $\text{Mg}^0\lambda = 1,737$ and $\lambda = 1,827$ and $\text{Mg}^+\lambda = 1,239$. We have shown (above) that the ionizing flux in the cloud is similar to that observed in our Galaxy, so we can estimate the electron density by equating the Galactic ionizing rate of Mg^0 to the recombination rate of Mg^+ (ref. 12). We find $n_e \approx 0.02 \text{ cm}^{-3}$. We note that the electronic density cannot be much smaller than this value because, from the excitation of the $J = 4$ and 5 H_2 levels, we know that the ionizing flux is not smaller than the Galactic value.

Then the hydrogen density can be derived from the $N(\text{C}^{+*})/N(\text{C}^+)$ ratio. Unfortunately, the $\text{C}^+\lambda = 1,334$ absorption line is saturated and cannot be used directly to infer $N(\text{C}^+)$. We therefore consider that the carbon metallicity can be derived from the silicon metallicity with some correction. Indeed, it is known that the metallicity of the α -chain elements is enhanced compared to carbon by a factor of about two²⁴ when the mean metallicity is low

Table 1 Heavy elements from the damped Lyman- α system

Ion	$\log N (\text{cm}^{-2})$	$b (\text{km s}^{-1})$	$[Z/H]$	$[Z/H] - [Z/H]_{\odot}$
H^0	20.90 ± 0.10	—	—	—
C^0	13.86 ± 0.22	1.70 ± 0.10	—	—
C^{0*}	13.43 ± 0.07	—	—	—
C^{0**}	12.63 ± 0.22	—	—	—
C^{++}	≤ 14.00	—	—	—
Mg^0	13.19 ± 0.09	—	—	—
Mg^+	15.44 ± 0.09	—	-5.46 ± 0.13	-1.04 ± 0.13
Si^+	15.24 ± 0.11	—	-5.66 ± 0.19	-1.20 ± 0.10
Fe^+	14.68 ± 0.08	—	-6.22 ± 0.13	-1.73 ± 0.13

N is the column density and b is the Doppler width of the line; $b = (2kT/m + b_{\text{turb}}^2)^{1/2}$, where m is the atomic mass, T the temperature and b_{turb} the characteristic turbulent velocity of the gas. $[Z/H] = \log N(Z) - \log N(\text{H})$ is the metallicity and $[Z/H] - [Z/H]_{\odot}$ is the metallicity relative to the value measured in the Sun. We used the solar metallicities from ref. 13. All wavelengths and oscillator strengths are from refs 12 and 28.

Table 2 Fit results to rotational levels in the vibrational ground state of H_2

z_{abs}	Level	$\log N (\text{cm}^{-2})$	$b (\text{km s}^{-1})$
2.33735	$J = 0$	$3.60 \pm 0.34 \times 10^{15}$	24.07
	$J = 1$	$3.34 \pm 0.40 \times 10^{15}$	24.07 ± 10.20
2.33754	$J = 0$	$4.10 \pm 0.48 \times 10^{15}$	14.09
	$J = 1$	$2.65 \pm 0.44 \times 10^{15}$	14.09 ± 4.57
2.33771	$J = 0$	$2.30 \pm 1.00 \times 10^{16}$	4.62 ± 0.36
	$J = 1$	$8.30 \pm 2.60 \times 10^{16}$	4.62
	$J = 2$	$1.59 \pm 0.27 \times 10^{16}$	4.62
	$J = 3$	$9.39 \pm 1.10 \times 10^{15}$	4.62
	$J = 4$	$4.42 \pm 0.80 \times 10^{14}$	4.62
	$J = 5$	$4.64 \pm 0.77 \times 10^{14}$	4.62
2.33793	$J = 0$	$1.00 \pm 0.04 \times 10^{15}$	21.74
	$J = 1$	$6.69 \pm 0.40 \times 10^{15}$	21.74 ± 1.40

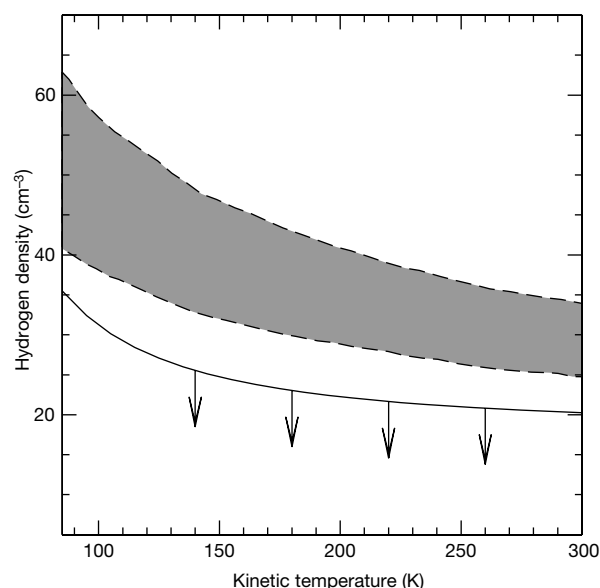


Figure 3 The hydrogen density as a function of kinetic temperature in the $z_{\text{abs}} = 2.33771$ cloud. The continuous curve with downward arrows shows the strict upper limit on the hydrogen density estimated from the fine-structure excitation of C^+ . The shaded area gives the 2σ range for the hydrogen density required to explain the population ratios in the fine-structure levels of C^0 , assuming there is no cosmic background radiation. All other excitation processes (collisions by electrons and hydrogen atoms, ultraviolet pumping) have been taken into account. It is apparent that the observed density is too small to explain the C^0 excitation, demonstrating that the cosmic microwave background radiation (CMBR) exists at $z = 2.33771$.

(that is, ≤ -1.0). To be conservative, we consider that $Z(C) = Z(\text{Si})/2$ and that silicon is not depleted into dust. We note that both assumptions maximize the C^{+*}/C^+ ratio and, as a consequence, maximize the derived hydrogen density, n_H . In addition, the very small C^0/C^+ value shows that the ionization correction is negligible. Also, the absorptions due to Si^{3+} and C^{3+} are unusually weak in this damped system. The fit of the $C^{+*}\lambda = 1,335.6$ and $\lambda = 1,335.7$ lines is shown in Fig. 2. The observed C^{+*} profile suggests the presence of possible extra components. Our measured column density is therefore an upper limit which will become an upper limit on the hydrogen density. The strongest constraint comes from the $\lambda = 1,335.6$ line in the blue wing of the profile.

Results on the determination of n_H are presented in Fig. 3. For a kinetic temperature in the range $85 < T < 285$ K, as suggested from the H_2 absorption, we find $20 < n_H < 35 \text{ cm}^{-3}$ (solid curve on Fig. 3). All the above assumptions maximize the hydrogen density, which can therefore be considered for each temperature as a conservative upper limit.

Cosmic microwave background radiation temperature

The fine-structure levels of C^0 can be populated by several processes, mainly collisions with hydrogen atoms and electrons, and pumping due to the local ultraviolet radiation and to the CMBR. The different contributions can be estimated once the temperature and the particle density are known. We have derived above $85 < T < 285$ K, $20 < n_H < 35 \text{ cm}^{-3}$ and $n_e/n_H \approx 0.001$, where n_e is the electron density.

Following ref. 5 and using cross-sections given in refs 22 and 23, we investigate the fine-structure excitation of C^0 , keeping the particle density and the temperature within the ranges we determined. For a radiation field similar to that of the Milky Way²⁴, the ultraviolet pumping rate from the ground state to the excited states of C^0 is $7.55 \times 10^{-10} \text{ s}^{-1}$. The corresponding value for the hydrogen

collisional rate in the density and temperature ranges considered here is of the order of $\sim 1.2\text{--}1.4 \times 10^{-8} \text{ s}^{-1}$. Thus the contribution due to ultraviolet pumping is negligible. Keenan *et al.*²⁵ noticed that collisions with electrons are unimportant for kinetic temperatures in the range 100–500 K and a ratio of electron density to hydrogen density that is less than 0.01. Collisions with electrons can therefore be neglected.

The observed 2σ range for $N(C^{0*})/N(C^0)$ and $N(C^{0**})/N(C^0)$ are, respectively, 0.28–0.48 and 0.020–0.060. Assuming collision by hydrogen atoms is the only process at work, we estimate the range of hydrogen density needed to explain the C^0 populations. This density range is shown as a shaded region in Fig. 3. It is apparent that the observed upper limit on the hydrogen density derived above is not sufficient to populate the excited fine-structure levels of C^0 and that an additional source of excitation is needed. The only process we are left with is direct pumping due to photons from the relic background radiation.

We estimate the allowed range for the CMBR temperature as a function of kinetic temperature, after taking into account all previously discussed processes. The allowed area in the $T_{\text{CMBR}}-T$ plane is shown as the shaded region in Fig. 4. This region is obtained using the upper limit on the hydrogen density derived above and the 2σ ranges of the C^{0*}/C^0 and C^{0**}/C^0 ratios. Thus the lower T_{CMBR} border of this area gives a stringent lower limit on the CMBR temperature at $z_{\text{abs}} = 2.33771$.

We note that an upper limit on T_{CMBR} is obtained by assuming that the CMBR is the only excitation process at work. This limit is shown as a dotted line in Fig. 4.

The presence of a relic radiation field at any redshift is more or less generally accepted in the literature. But its reality has been proved observationally only in the local universe through direct measurements^{3,26} and the Sunyaev & Zeldovich effect²⁷. Our measurement demonstrates for the first time, to our knowledge, that the

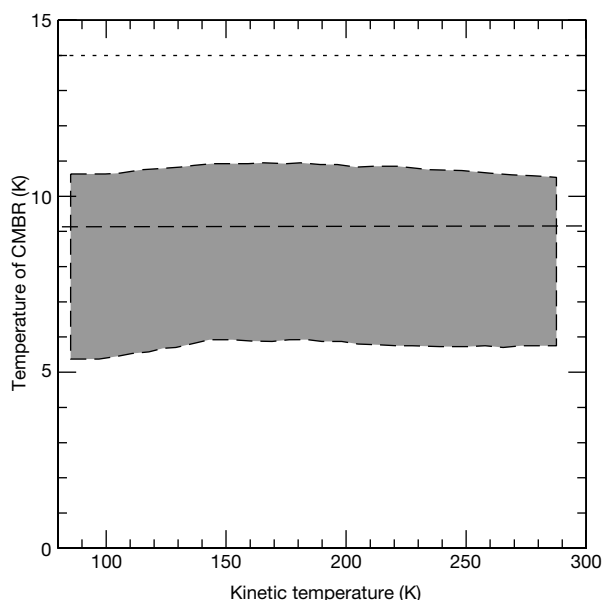


Figure 4 Cosmic microwave background temperature as a function of kinetic temperature of the gas. For a given temperature, we obtain an upper limit of the hydrogen density from the C^+ fine-structure population. This is used to derive the contribution of hydrogen collisions to the C^0 fine-structure excitation. The excess is used to determine the range for the CMBR temperature. The shaded region gives the 2σ range of the CMBR radiation temperature allowed by the observed population ratios of neutral carbon fine-structure levels. The horizontal dotted line is the upper limit on T_{CMBR} if CMBR is assumed to be the only excitation process. The dashed line is the predicted temperature at $z = 2.33771$ in the standard Big Bang model. This demonstrates the presence of a background radiation with a temperature at least twice that measured in the local universe, $T_{\text{CMBR},0} = 2.726$ K.

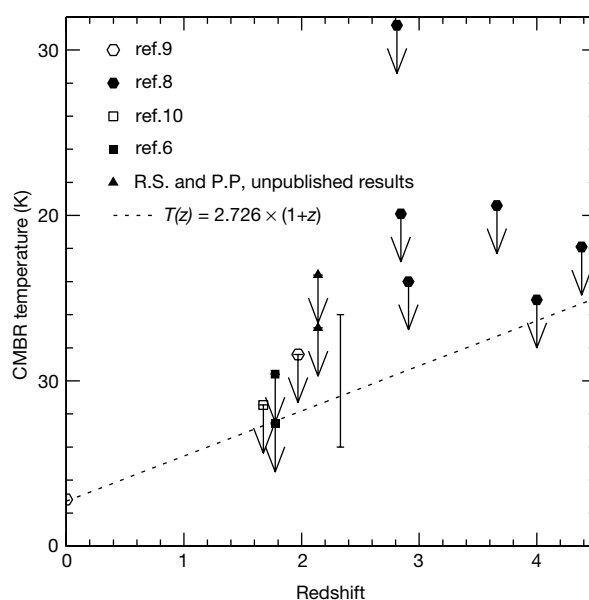


Figure 5 Measurements of the cosmic microwave background radiation temperature at various redshifts. The point at $z = 0$ shows the result of the Cosmic Background Explorer (COBE) determination³, $T_{\text{CMBR}}(0) = 2.726 \pm 0.010$ K. Upper limits are previous measurements^{3,8–10} using the same techniques as we did. We also include our two new unpublished upper limits at $z = 2.1394$ along the line of sight toward Tololo 1037–270. The measurement from this work, $6.0 < T_{\text{CMBR}} < 14.0$ K at $z = 2.33771$, is indicated by a vertical bar. The dashed line is the prediction from the hot Big Bang, $T_{\text{CMBR}}(z) = T_{\text{CMBR}}(0) \times (1 + z)$.

cosmic radiation exists at earlier times and has a higher temperature than today. In the standard Big Bang cosmology the cosmic microwave background is a relic radiation left over from an early hot phase. In such a model the radiation evolves adiabatically in the expanding universe and the temperature at any redshift is simply $T(z) = T(z=0)(1+z)$. We summarize all the available upper limits in Fig. 5, together with our measurement. The prediction for the adiabatic expansion is shown as a dotted line. These upper limits and our measurements are consistent with the predictions of the standard model. Similar analysis over a large redshift range will provide a direct, model-independent measure of the evolution of the cosmic microwave background radiation. □

Received 7 August; accepted 3 November 2000.

1. Alpher, R. A., Bethe, H. A. & Gamov, G. Evolution of chemical elements. *Phys. Rev.* **73**, 803–804 (1948).
2. Penzias, A. A. & Wilson, R. A measurement of excess antenna temperature at 4080 Mc/s. *Astrophys. J.* **142**, 419–421 (1965).
3. Mather, J. C. *et al.* Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument. *Astrophys. J.* **420**, 439–444 (1994).
4. Bahcall, J. N., Joss, P. C. & Lynds, R. On the temperature of the microwave background radiation at a large redshift. *Astrophys. J.* **182**, L95–L98 (1973).
5. Meyer, D. M., Black, J. H., Chaffee, F. H., Foltz, C. B. & York, D. G. An upper limit on the microwave background temperature at $z = 1.776$. *Astrophys. J.* **308**, L37–L41 (1986).
6. Songaila, A. *et al.* Measurement of the microwave background temperature at a redshift of 1.776. *Nature* **371**, 43–45 (1994).
7. Songaila, A., Cowie, L. L., Hogan, C. & Rugers, M. Deuterium abundance and background radiation temperature in high redshift primordial clouds. *Nature* **368**, 599–604 (1994).
8. Lu, L., Wargent, W. L. W., Womble, D. S. & Barlow, T. A. Abundances at high redshifts: The chemical enrichment history of damped Lyman- α galaxies. *Astrophys. J. Suppl. Ser.* **107**, 475–520 (1996).
9. Ge, J., Bechtold, J. & Black, J. H. A new measurement of the cosmic microwave background radiation temperature at $z = 1.97$. *Astrophys. J.* **474**, 67–73 (1997).
10. Roth, K. C. & Bauer, J. M. The $z = 1.6748$ C I absorber toward PKS 1756+237. *Astrophys. J.* **515**, L57–L60 (1999).
11. D'odorico, S. *et al.* Performance of UVES, the echelle spectrograph for the ESO VLT and highlights of the first observations of stars and quasars. *Proc. SPIE* **4005**, 121–130 (2000).
12. Welty, D. E. *et al.* The diffuse interstellar cloud toward 23 Orionis. *Astrophys. J. Suppl. Ser.* **124**, 465–501 (1999).
13. Savage, B. D. & Sembach, K. R. Interstellar abundances from absorption-line observations with the Hubble space telescope. *Annu. Rev. Astron. Astrophys.* **34**, 279–330 (1996).
14. Pettini, M., Smith, L. J., King, D. L. & Hunstead, R. W. The metallicity of high-redshift galaxies: the abundance of zinc in 34 damped Lyman- α systems from $z = 0.7$ to 3.4. *Astrophys. J.* **486**, 665–680 (1997).
15. Ge, J. & Bechtold, J. in *Highly Redshifted Radio Lines* (eds Carilli, C. L., Radford, S. J. E., Menten, K. M. & Langston, G. I.) 121–131 (ASP Conf. Series Vol. 156, 1999).
16. Shull, J. M. *et al.* FUSE observations of diffuse interstellar molecular hydrogen. *Astrophys. J.* **538**, L73–L76 (2000).
17. Jura, M. Interstellar clouds containing optically thick H₂. *Astrophys. J.* **197**, 581–586 (1975).
18. Dalgarno, A. & Wright, E. L. Infrared emissivities of H₂ and HD. *Astrophys. J.* **174**, L49–L51 (1972).
19. Srianand, R. & Petitjean, P. Molecules in the $z_{\text{abs}} = 2.8112$ damped system toward PKS 0528-250. *Astron. Astrophys.* **335**, 33–40 (1998).
20. Bahcall, J. N. & Wolf, R. A. Fine-structure transitions. *Astrophys. J.* **152**, 701–729 (1968).
21. Timmes, F. X., Lauroesch, J. J. & Truran, J. W. Abundance histories of QSO absorption systems. *Astrophys. J.* **518**, 468–476 (1995).
22. Nussbaumer, H. & Rusca, C. Forbidden transitions in the C I sequence. *Astron. Astrophys.* **72**, 129–133 (1979).
23. Launay, J. M. & Roueff, E. Fine structure excitation of carbon and oxygen by atomic hydrogen impact. *Astron. Astrophys.* **56**, 289–292 (1977).
24. Jenkins, E. B. & Shaya, E. J. A survey of interstellar C I insight on carbon abundances UV grain albedoes, and pressures in the interstellar medium. *Astrophys. J.* **231**, 55–72 (1979).
25. Keenan, F. P., Lennon, D. J., Johnson, C. T. & Kingston, A. E. Fine structure population for the 2P ground states of C II. *Mon. Not. R. Astron. Soc.* **220**, 571–576 (1986).
26. Roth, K. C. & Meyer, D. M. Cyanogen excitation in diffuse interstellar cloud. *Astrophys. J.* **441**, 129–143 (1995).
27. Sunyaev, R. A. & Zeldovich, Ya. B. The velocity of clusters of galaxies relative to the microwave background—The possibility of its measurement. *Mon. Not. R. Astron. Soc.* **190**, 413–420 (1980).
28. Morton, D. C. Atomic data for resonance absorption lines. I—Wavelengths longward of the Lyman limit. *Astrophys. J. Suppl. Ser.* **77**, 119–202 (1991).

Acknowledgements

The observations presented here have been obtained using the Ultra-violet and Visible Echelle Spectrograph mounted on the 8.2-m KUEYEN telescope operated by the European Southern Observatory at Paranal, Chile. P.P. thanks A. Kaufer and M. Chadid for their kind and efficient assistance at the telescope and IUCAA for hospitality during the time this work was being done. We thank T. Padmanabhan for useful comments. We gratefully acknowledge support from the Indo-French Centre for the Promotion of Advanced Research (Centre Franco-Indien pour la Promotion de la Recherche Avancée).

Correspondence and requests for materials should be addressed to R.S. (e-mail: anand@iucaa.ernet.in).

Stargazin regulates synaptic targeting of AMPA receptors by two distinct mechanisms

Lu Chen^{*†}, Dane M. Chetkovich^{†‡§}, Ronald S. Petralia^{||}, Neal T. Sweeney[‡], Yoshimi Kawasaki[‡], Robert J. Wenthold^{||}, David S. Bredt[‡] & Roger A. Nicoll^{*‡}

Departments of ^{*} Cellular and Molecular Pharmacology, [‡] Physiology, and [§] Neurology, University of California, San Francisco, California 94143, USA

^{||} Laboratory of Neurochemistry, NIDCD, National Institutes of Health, Bethesda, Maryland 20892, USA

[†] These authors contributed equally to this work

Stargazer, an ataxic and epileptic mutant mouse, lacks functional AMPA (α -amino-3-hydroxyl-5-methyl-4-isoxazolepropionate) receptors on cerebellar granule cells. Stargazin, the mutated protein, interacts with both AMPA receptor subunits and synaptic PDZ proteins, such as PSD-95. The interaction of stargazin with AMPA receptor subunits is essential for delivering functional receptors to the surface membrane of granule cells, whereas its binding with PSD-95 and related PDZ proteins through a carboxy-terminal PDZ-binding domain is required for targeting the AMPA receptor to synapses. Expression of a mutant stargazin lacking the PDZ-binding domain in hippocampal pyramidal cells disrupts synaptic AMPA receptors, indicating that stargazin-like mechanisms for targeting AMPA receptors may be widespread in the central nervous system.

Excitatory synapses in the central nervous system (CNS) release glutamate onto a number of receptor subtypes. The principal ionotropic glutamate receptors include AMPA receptors (AMPARs) and NMDARs (*N*-methyl-D-aspartate receptors). AMPARs mediate moment-to-moment signalling, whereas NMDARs initiate synaptic plasticity. Recent studies emphasize remarkable differences in synaptic expression of these receptors. NMDARs are relatively fixed components of the postsynaptic density (PSD), whereas AMPARs are more loosely associated and their density at synapses is tightly controlled by neuronal activity^{1–8}. That the number of AMPARs at the synapse is regulated independently of NMDARs raises intriguing questions concerning mechanisms involved in synaptic targeting of glutamate receptors and the role that this plasticity plays in learning and memory.

This independent regulation of synaptic AMPARs versus NMDARs is clearly manifested in the stargazer mutant mouse, which exhibits seizures and cerebellar ataxia. Among the defects identified in the cerebellum is the lack of functional AMPARs on granule cells^{9,10}. The defective protein stargazin was recently identified as a relative of the γ -1 subunit of the skeletal muscle calcium channel¹¹. More recently, a family of γ -like subunits has been identified, each with distinct expression patterns in the brain^{12,13}. Why stargazin is necessary for functional AMPARs in cerebellar granule cells has remained mysterious and is the focus of this study.

stg/stg granule cells lack AMPAR synaptic currents

We first determined whether the defect in AMPARs in the stargazer mouse results from abnormal cerebellar circuitry or alternatively is an autonomous granule cell defect. We therefore evaluated AMPARs in granule cell cultures, which lack the normal excitatory mossy fibre input, as well as Purkinje cells, the primary target for granule cells. In *+stg* cultures, spontaneous rapid inward currents are routinely recorded from individual granule cells, presumably generated by synapses between granule cells (Fig. 1a). These events are largely mediated by the action potential dependent release of glutamate onto AMPARs, as their frequency is greatly reduced by tetrodotoxin (TTX) and they are abolished by the AMPAR antagonist CNQX (data not shown). In striking contrast, *stg/stg* neurons exhibit essentially no spontaneous activity (Fig. 1b). The average

size of events in *+stg* cells was 18.3 ± 5.1 pA ($n = 9$), whereas the average size in *stg/stg* cells was 4.8 ± 0.4 pA ($n = 18$), close to the threshold value of 4 pA used to detect these events (Fig. 1c, see broken line in graph). Furthermore, the frequency of events was 1.7 ± 0.7 Hz ($n = 9$) in *+stg* and 0.1 ± 0.1 Hz ($n = 18$) ($P < 0.01$) in *stg/stg* (data not shown).

The *stg/stg* neurons clearly form excitatory synaptic connections as NMDAR currents can be recorded when Mg^{2+} is removed from, and CNQX and glycine are added to, the solution (compare Fig. 1d and e). As expected for NMDARs, these currents are considerably slower than those mediated by AMPAR currents and are blocked by the NMDAR antagonist AP5 (*D*(-)-amino-5-phosphonovaleric acid (data not shown). Figure 1f shows that there was no difference in the amplitude of NMDAR-mediated events in *+stg* cells (15.4 ± 1.2 pA, $n = 20$) and *stg/stg* cells (14.1 ± 2.4 pA, $n = 9$), or in the frequency of these events (*+stg* = 0.5 ± 0.1 Hz, $n = 20$; *stg/stg* = 0.4 ± 0.1 Hz, $n = 9$; $P > 0.1$).

Localization of AMPAR subunits in granule cells

In cerebellar cultures from *+stg* mouse, the AMPAR subunit GluR4 forms discrete synaptic puncta that colocalize with the presynaptic marker synaptophysin1 (Fig. 1g). By contrast, few synaptic GluR4 puncta are evident in *stg/stg* cells. Despite lacking synaptic GluR4 puncta, the cultured *stg/stg* granule cells are decorated with presynaptic synaptophysin1 puncta. To assess the synaptic localization of AMPAR subunits at mossy fibre synapses in the intact cerebellar glomeruli, we used post embedding immunogold electron-microscopic analysis. Granule cell synapses in *stg/stg* cerebellum are virtually devoid of GluR2/3 labelling, whereas these synapses in *+stg* are labelled abundantly (Fig. 2a, b). Similar results were obtained with an antibody to the GluR4 subunit; most synapses in the *+stg* mouse are GluR4-positive (Fig. 2d), whereas synapses in the *stg/stg* mouse are rarely labelled (Fig. 2c). A 'blind' quantitative analysis showed roughly a 10-fold decrease in the number of GluR2/3-reactive particles at *stg/stg* synapses (Table 1).

In contrast to the defect in synaptic GluR 2/3 labelling, the NMDAR subunit NR1 labelling was increased in the *stg/stg* mutant (gold particles per synapse: *+stg*, 0.40; *stg/stg*, 0.84; $P < 1 \times 10^{-5}$). The cytoplasmic GluR2/3 labelling was low in both *+stg*

and *stg/stg* granule cells, making quantification difficult, but no obvious difference was noted. The presence of cytoplasmic staining indicates that AMPARs are present in the *stg/stg* granule cells, in agreement with western blot analysis¹⁰. Finally, there was no obvious difference in the pre- or postsynaptic morphology of the mossy fibre to granule cell synapse between the *stg/stg* mice and *+stg* mice. For example, the length of postsynaptic densities averaged 149 nm for *stg/stg* ($n = 227$) and 143 nm for *+stg* ($n = 242$) ($P > 0.1$).

Calcium currents are normal in *stg/stg* granule cells

On the basis of its weak homology to the γ -1 calcium channel subunit¹¹, we looked for altered calcium currents in *stg/stg* granule cells. In cultured cerebellar granule cells, voltage-dependent calcium currents can be well clamped and are largely mediated by P/Q-, L- and N-type channels¹⁴. There was no difference in the peak amplitude (control = 83.9 ± 10.3 pA, $n = 11$; *stg/stg* = 88.7 ± 9.3 , $n = 9$; $P > 0.5$) (data not shown), the activation curve for the calcium currents (Fig. 3a), or the steady-state inactivation (Fig. 3b) between *stg/stg* and *+stg* cells. Together with the absence of AMPAR currents, these data make it most unlikely that the primary defect in *stg/stg* granule cells is altered calcium channel function.

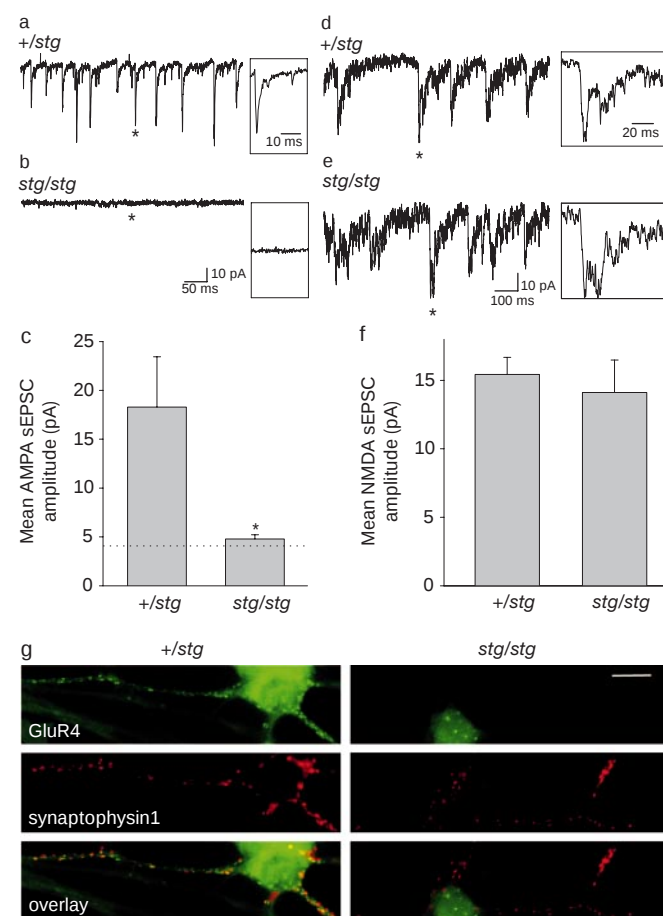


Figure 1 Functional AMPARs are absent at *stg/stg* granule cell synapses in culture. **a, b**, Spontaneous AMPAR EPSCs are abundant in *+stg* granule cells, but are absent in *stg/stg* granule cells. Asterisk, individual event displayed in the inset on the right. **c**, Mean amplitude of AMPAR EPSCs from *+stg* and *stg/stg* granule cells ($P < 0.0005$). The detection threshold, 4 pA (dashed line), was assigned to cells in the *stg/stg* group that did not show any synaptic AMPAR responses during the recording period. **d, e**, Spontaneous NMDAR EPSCs from *+stg* and *stg/stg* granule cells in the presence of CNQX (10 μ M), glycine (20 μ M) and 0 Mg^{2+} . **f**, Mean amplitude of NMDAR EPSCs from *+stg* and *stg/stg* granule cells ($P > 0.5$). **g**, Immunostaining of GluR4 and synaptophysin in methanol-fixed 8-DIV granule cell cultures. *+stg* cells, in contrast to *stg/stg* cells, exhibit dendritic GluR4 puncta that colocalized with synaptophysin. Scale bar, 5 μ m.

Stargazin interacts with AMPARs and PDZ proteins

As stargazin granule cells lack AMPAR currents, we next determined whether stargazin associates with GluR subunits. When co-transfected into COS cells, stargazin and GluR4 co-immunoprecipitate (Fig. 4a). Notably, stargazin also interacts with co-transfected GluR1 or 2 subunits (Fig. 4b). These stargazin interactions with GluR subunits appear specific, as stargazin does not bind to NMDAR (NR1-4b) (Fig. 4c) or Kv1.4 (data not shown). In brain homogenates, stargazin is enriched in Triton X-100 insoluble postsynaptic density fractions together with GluR4, NR1 and postsynaptic density-95 (PSD-95) (Fig. 4g). Although we were unable to co-immunoprecipitate stargazin with GluR from brain extracts (data not shown), the harsh conditions required for solubilization of the PSD disrupt many protein complexes and often preclude detection of interactions^{15,16}.

Stargazin is predicted to contain four putative transmembrane domains with intracellular amino- and carboxy-terminal tails¹¹. We noticed that its C-terminal region contains a type I PDZ-binding site for association with certain synaptic PDZ proteins such as PSD-95. We found that stargazin interacts with PSD-95 (Fig. 4d) and SAP-97 (Fig. 4e), as well as with PSD-93 and SAP-102 (data not shown).

Table 1 Summary of immunogold labelling for GluR2/3 at cerebellar glomeruli

	<i>+stg</i>		<i>stg/stg</i>	
	No. 1	No. 2	No. 1	No. 2
Synapses examined	123	191	140	150
Per cent synapses labelled*	79%	82%	18%	13%
Particles/synapse*	2	1.94	0.26	0.21
Particles/labelled synapse*	2.54	2.36	1.48	1.55

*The percentage of labelled synapse, the average number of particles per synapse and the average number of particles per labelled synapse were significantly different (t -test) between *+stg* and *stg/stg*. No significant difference was found between two animals (no. 1 and 2) within each group.

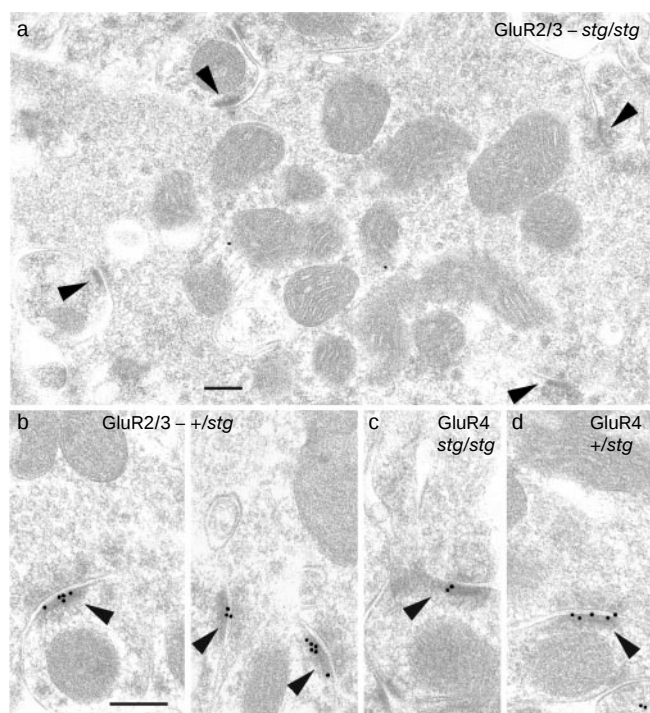


Figure 2 AMPAR immunogold labelling of cerebellar mossy fibre-granule cell synapses. The examples illustrate the low labelling of *stg/stg* synapses compared with that of *+stg* synapses. **a**, No GluR2/3-positive gold particles are present at the four asymmetric synapses (arrow heads) made by a single mossy fibre terminal in *stg/stg* cerebellum. **b**, Typical synaptic labelling with the GluR2/3 antibody in *+stg* mice. **c, d**, Examples of synapses in the *stg/stg* (**c**) and *+stg* (**d**) mouse labelled for GluR4. Scale bar, 0.2 μ m.

shown). Deleting the final four amino acids of stargazin (stargazin Δ C) disrupts interaction with PSD-95 (data not shown), although stargazin Δ C does retain binding to GluR4 (Fig. 4f).

Because PSD-95 can mediate clustering of ion channels¹⁷, we wondered whether its interactions with stargazin might cluster GluRs in heterologous cells. Co-expression of GluR4 with PSD-95 (Fig. 4h, top row), or with stargazin (middle row) results in diffuse, overlapping distributions for these proteins, which generally resemble the localization of GluR4 expressed alone (data not shown). However, transfecting the three together causes a remarkable redistribution of the proteins to patch-like clusters (Fig. 4h, bottom row). This clustering requires interaction of stargazin with the PDZ domains from PSD-95, as clustering is not observed in co-transfections with stargazin Δ C (data not shown). The stargazin/PSD-95-induced clusters of GluR4 occur at the cell surface, as they are labelled in non-permeabilized cells by an antibody to the extracellular green fluorescent protein (GFP) tag on GluR4 (data not shown).

AMPA responses to glutamate in *stg/stg* granule cells

The absence of spontaneous activity in *stg/stg* granule cells suggests

that stargazin mediates synaptic targeting, but does not address whether extrasynaptic receptors are altered. We therefore analysed granule cell responses to rapid application of glutamate. Responses were first recorded at a membrane potential of -80 mV, which isolates AMPAR responses, and these were compared to responses evoked at $+60$ mV, where both AMPAR and NMDAR currents are recorded (Fig. 5a, b). In *+stg* neurons, a considerable current is generated at -80 mV (77 ± 8 pA, $n = 7$), and this is largely abolished by NBQX (10 ± 0.3 pA, $n = 4$), confirming that at -80 mV this current is carried by AMPARs. In striking contrast, *stg/stg* neurons show very little current at -80 mV (9 ± 3 pA, $n = 5$). The small NBQX resistant component, similar in size to that recorded in *stg/stg* granule cells, is presumably generated by NMDARs, which contain the NR2C subunit in granule cells and have an incomplete Mg^{2+} blockade at negative potential¹⁸. At $+60$ mV, a large outward current is recorded in *+stg* neurons (177 ± 32 pA, $n = 7$), which is not significantly different from the current recorded in *stg/stg* neurons (257 ± 57 pA, $n = 5$). NBQX has little effect on the *+stg* responses recorded at $+60$ mV (162 ± 22 pA, $n = 4$), because the low concentration of glutamate used (100μ M) more effectively activates NMDARs compared with AMPARs. These results illustrate the complete and selective absence of functional AMPARs in *stg/stg* granule cells.

Stargazin rescues AMPAR responses in *stg/stg* cells

Transfecting stargazin in *stg/stg* neurons restores synaptic AMPAR function as stargazin-GFP expressing cells display a large number of spontaneous inward currents (15.8 ± 2.7 pA, $n = 7$) as compared with GFP-expressing cells (4.8 ± 0.9 pA, $n = 3$) or untransfected cells (4.8 ± 0.4 pA, $n = 19$) (Fig. 6a, b). Stargazin transfection increased the frequency of events from 0.1 ± 0.1 Hz ($n = 19$) to 1.3 ± 0.3 Hz ($n = 7$) ($P < 0.0005$) (data not shown). By contrast, expression of stargazin Δ C failed to restore synaptic responses (5.0 ± 0.7 pA, $n = 8$).

We also found that glutamate-evoked responses were restored by stargazin transfection. First, to assess any possible role for voltage-dependent calcium channels (VDCCs) in the AMPAR phenotype, we transfected stargazin into 5-DIV (days *in vitro*) *stg/stg* cultures that had all VDCCs blocked by 10μ M cadmium from the time of transfection. Despite the continuous absence of calcium currents (data not shown), glutamate-evoked responses were restored 24 h after transfection (control, 4.8 ± 1.7 pA, $n = 3$; transfected, 44.4 ± 7.0 pA, $n = 6$; $P < 0.005$). Second, we examined the surface-receptor response in stargazin Δ C-transfected *stg/stg* cultures. Unexpectedly, stargazin Δ C fully restores responses in *stg/stg* cells at -80 mV (*stg/stg* = 14.6 ± 4.4 pA, $n = 5$; versus *stg/stg* plus stargazin Δ C = 205.7 ± 56.6 pA, $n = 7$), but had no significant effect on responses recorded at $+60$ mV (Fig. 6c, d). In wild-type cells, stargazin Δ C has little effect on AMPAR responses (-80 mV, untransfected: 104.3 ± 10.1 pA, $n = 26$; transfected: 141 ± 36.0 pA, $n = 10$; $P > 0.1$) or on the NMDAR responses ($+60$ mV, untransfected, 205.1 ± 25.5 pA, $n = 18$; transfected, 306.0 ± 91.1 pA, $n = 10$; $P > 0.1$) to glutamate applications (Fig. 6e). Although stargazin Δ C has no effect on the AMPAR response to glutamate application in wild-type neurons (Fig. 6e), it markedly reduced the amplitude (control = 14.4 ± 0.7 pA, $n = 22$; stargazin Δ C = 7.0 ± 0.7 pA, $n = 18$; $P < 1 \times 10^{-7}$) and frequency (control = 1.1 ± 0.2 Hz, $n = 22$; stargazin Δ C = 0.4 ± 0.2 Hz, $n = 18$; $P < 0.05$) of spontaneous AMPAR synaptic events (Fig. 6f), indicating that stargazin Δ C disrupts synaptic localization of AMPARs.

Stargazin Δ C downregulates hippocampal AMPAR EPSCs

Why is the AMPAR defect in stargazer mice restricted to cerebellar granule cells^{9,10}? Perhaps the same mechanism applies to all excitatory synapses, but stargazin isoforms are expressed in other regions of the brain. Indeed, studies have shown that a family of three stargazin related proteins is expressed throughout the CNS¹³. *In situ*

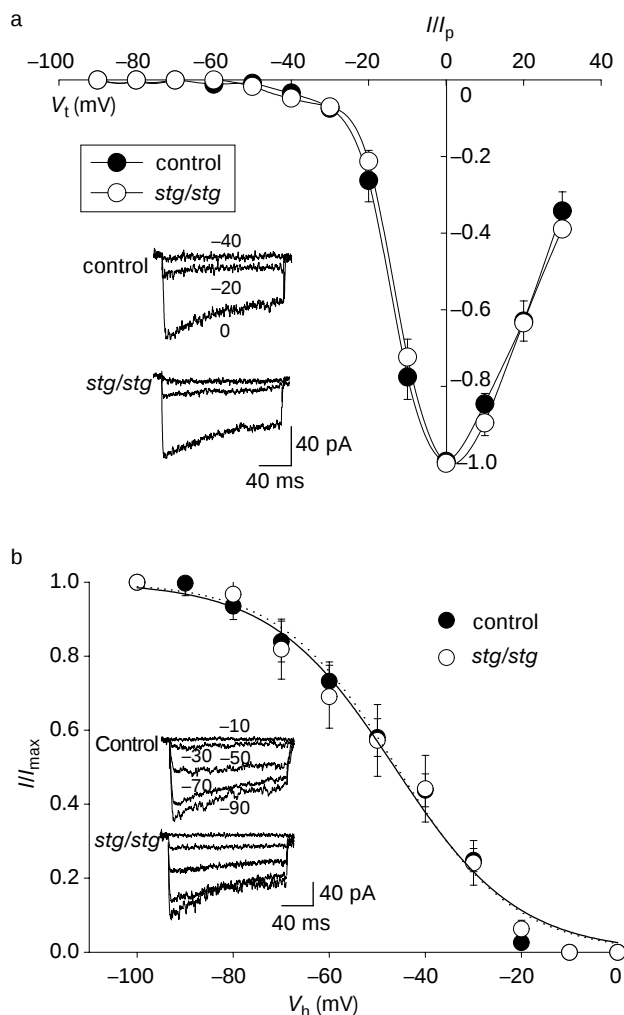


Figure 3 Ca^{2+} currents recorded from *+stg* and *stg/stg* granule cells. **a**, Activation of whole-cell Ca^{2+} currents ($n = 8$ for each group). Inset, activation of Ca^{2+} current with depolarizing test potentials ($V_t = -40, -20$ and 0 mV). **b**, Steady-state inactivation of Ca^{2+} currents. Individual currents in each cell were evoked by a 150-ms voltage pulse to 0 mV from a 10-s inactivating prepulse ($V_h = -100$ to 0 mV) ($n = 11$ for each group). Data were fitted with a single Boltzmann function, $I/I_{max} = 1/[1 + \exp((V - V_{1/2})/k)]$. *+stg*, $V_{1/2} = -46.0 \pm 1.8$ mV and $k = 12.3 \pm 1.4$ mV; *stg/stg*, $V_{1/2} = -46.4 \pm 2.0$ mV and $k = 12.8 \pm 1.6$ mV.

hybridization shows that stargazin (γ -2) is expressed in cerebellar granule cells, cerebellar Purkinje cells and hippocampal pyramidal cells (Fig. 7a, top). The stargazin related gene γ -3 is not expressed in cerebellar granule cells but does occur in cerebellar Purkinje cells and hippocampal pyramidal cells (Fig. 7a, bottom). We evaluated possible functions for stargazin proteins in forebrain by transfecting cultured hippocampal cells. Transfected stargazin has a punctate distribution in hippocampal neurons (Fig. 7b, left), whereas stargazin Δ C is diffuse (Fig. 7b, right). The stargazin clusters are synaptic as shown by colocalization with GluR1 (Fig. 7c), SV2 and synaptophysin1 (data not shown). Stargazin appears restricted to excitatory synapses, as it does not colocalize with inhibitory boutons stained for glutamic-acid decarboxylase 65 (GAD 65) (Fig. 7d).

We wondered whether stargazin Δ C might disrupt AMPAR EPSCs in hippocampal neurons by preventing interaction between AMPARs and other members of the stargazin family. Stargazin Δ C

had no effect on glutamate (1 μ M) induced currents at -80 mV (control = 413 ± 170 pA, $n = 11$; stargazin Δ C = 425 ± 109 pA, $n = 8$), or at $+60$ mV (control = 770 ± 174 pA, $n = 11$; stargazin Δ C = 700 ± 146 pA, $n = 8$) (data not shown). However, stargazin Δ C expression decreased markedly the amplitude of AMPAR miniature EPSCs (mEPSCs) (control = 20.2 ± 1.1 pA, $n = 15$; stargazin Δ C = 10.1 ± 1.0 pA, $n = 14$) when compared with neighbouring untransfected neurons (Fig. 7e, left). In addition, the frequency of mEPSCs was reduced greatly (control = 15.7 ± 4.2 Hz, $n = 15$; stargazin Δ C = 1.6 ± 0.9 Hz, $n = 14$) (Fig. 7e, right).

Discussion

This study defines an essential role for stargazin in synaptic expression of glutamate receptors in the CNS. Previous studies have focused on two possible roles for stargazin. Homology to the γ -1 subunit of the calcium channel has suggested that stargazin may be a

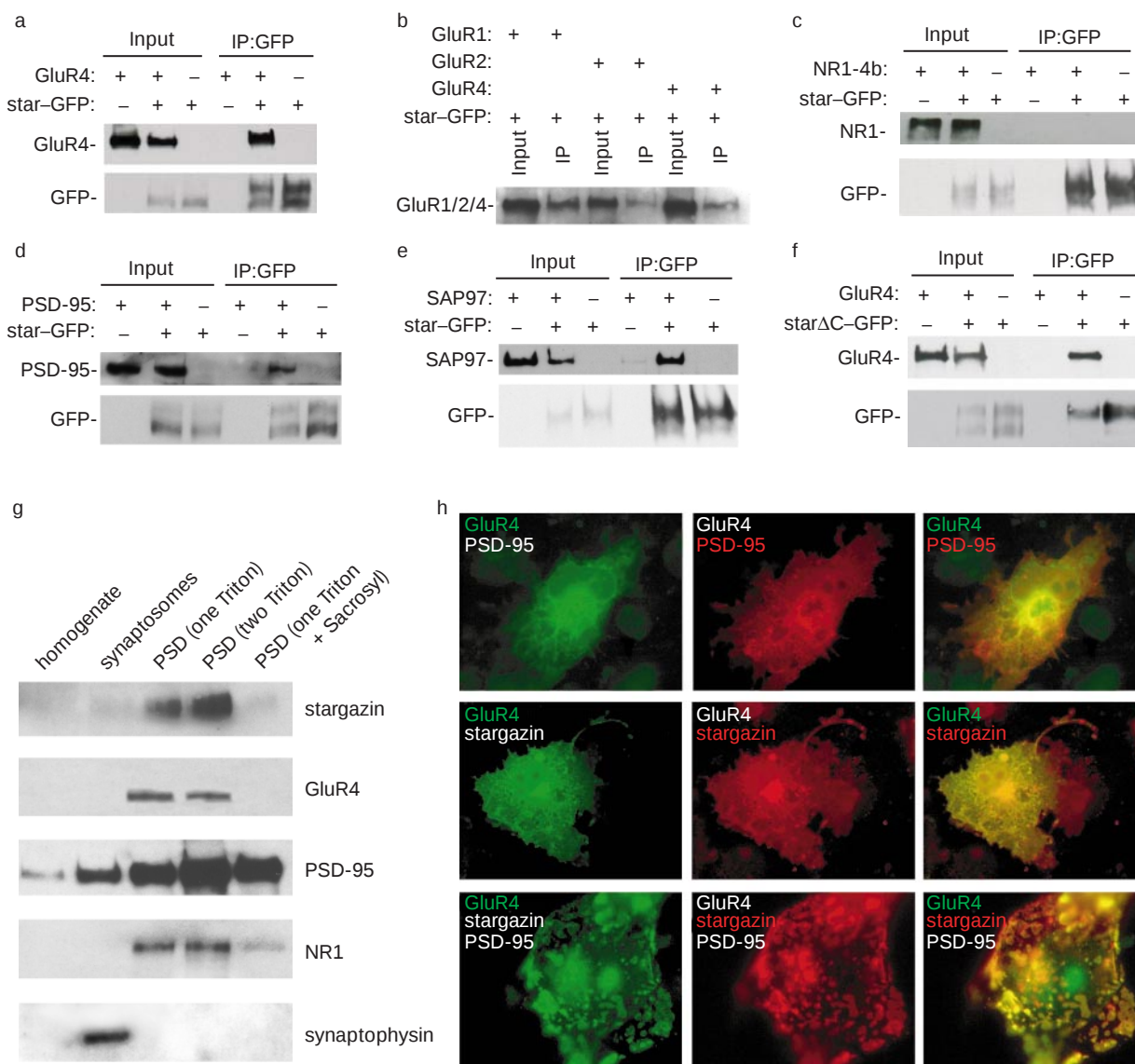


Figure 4 Interactions between Stargazin, GluR and PDZ proteins. **a**, **b**, Co-immunoprecipitation of stargazin and AMPAR subunits from COS cells. Cells were transfected with stargazin-GFP and GluR4 (**a**) or stargazin-GFP and GluR1, GluR2 or GluR4 (**b**). **c**, Stargazin fails to immunoprecipitate Myc-tagged NR1. **d**, **e**, Co-immunoprecipitation of stargazin with PSD-95 (**d**) and with SAP-97 (**e**). **f**, Deletion of the C-terminal PDZ-binding domain from stargazin (star Δ C-GFP) does not prevent the interaction of stargazin with GluR4. **g**, Stargazin is enriched in the PSD fraction resistant to

solubilization with Triton X-100. Antibodies against GluR4, PSD-95 and NR1 were used as controls for postsynaptic proteins, and an antibody against synaptophysin was used as a presynaptic control. **h**, COS cells were transfected with different combinations of GFP-GluR4, PSD-95 and stargazin-HA as indicated. After permeabilization and fixation, the distribution of transfected proteins was visualized by indirect immunofluorescence. The protein specifically visualized is indicated by coloured text.

calcium-channel subunit in the CNS¹¹. Alternatively, deficits in cerebellar function^{9,10} suggested that stargazin may regulate AMPARs. The homology of stargazin with the γ -1 subunit is weak (less than 25%), and the physiological effects of stargazin on calcium channels are modest^{11,13}. We were unable to detect any difference between the calcium channel currents recorded in *stg/stg* or *+stg* granule cells. Furthermore, AMPAR currents were rescued

despite blockade of voltage-dependent calcium channels with cadmium, indicating that stargazin effects on AMPAR function are independent of calcium channels.

Our work shows that stargazin has a direct interaction with AMPARs. Stargazin co-immunoprecipitates with GluR1, -2 and -4 in heterologous cells. Ultrastructural studies using immunogold labelling found a marked decrease in synaptic GluR2/3 and GluR4

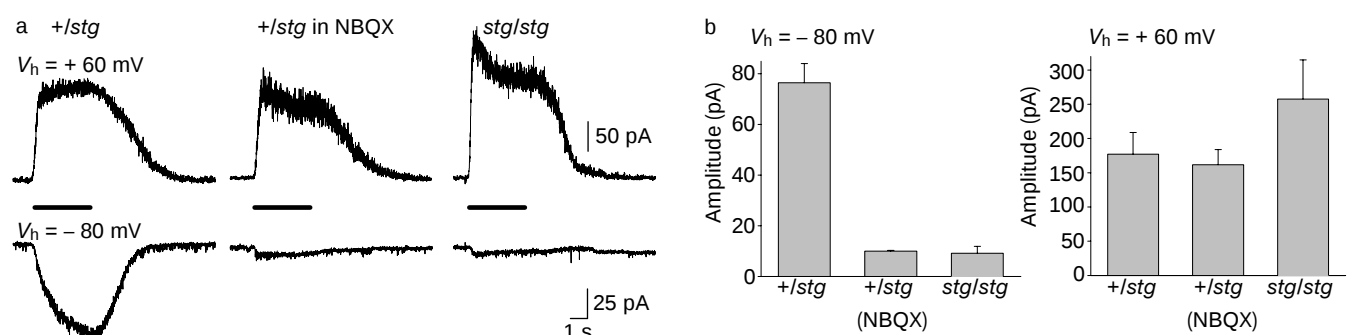


Figure 5 Surface AMPARs are absent in *stg/stg* granule cells. **a**, Examples of whole-cell currents ($V_h = -80$ mV or $+60$ mV) evoked by fast application of glutamate ($100 \mu\text{M}$) in the presence of TTX ($1 \mu\text{M}$) and glycine ($20 \mu\text{M}$). At -80 mV, glutamate evoked an inward current in a *+stg* neuron, which is largely absent in NBQX ($15 \mu\text{M}$) or in *stg/stg* neurons.

b, Summary of experiments described in **a**. *stg/stg* granule cells exhibited virtually no AMPAR generated currents (-80 mV), nor did the *+stg* cells in the presence of NBQX (*+stg* versus *stg/stg*, $P > 0.5$) although their NMDAR currents ($+60$ mV) were normal ($P > 0.1$).

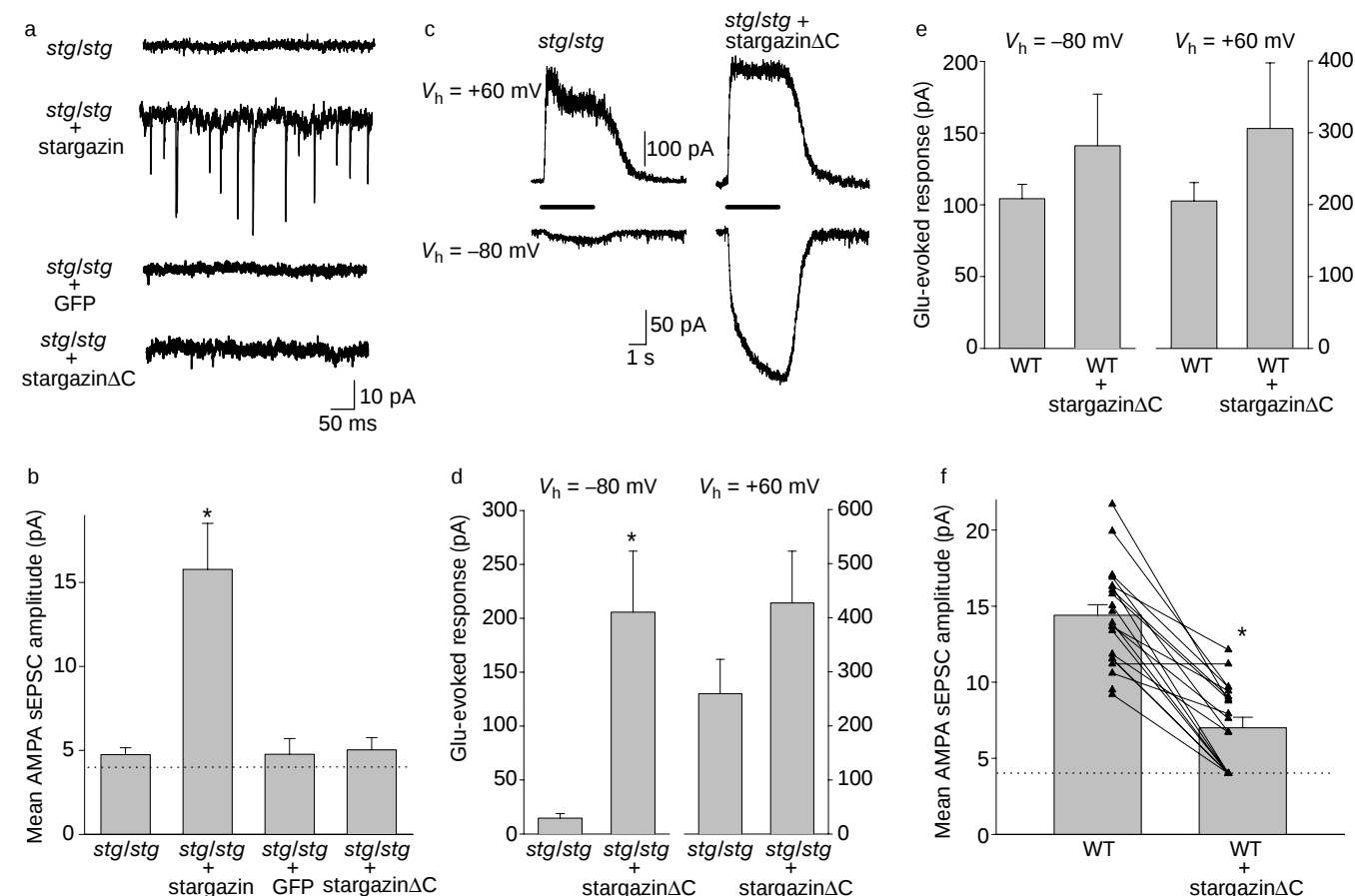


Figure 6 Stargazin rescues AMPAR responses in *stg/stg* granule cells. **a**, **b**, Transfection of stargazin-GFP, but not GFP or stargazin Δ C, restores synaptic AMPAR responses in *stg/stg* granule cells (asterisk, $P < 0.0001$). Dashed line in **b** indicates the detection threshold (4 pA): cells without synaptic events during the recording period were assigned a response amplitude of 4 pA. **c**, **d**, Stargazin Δ C restores glutamate-evoked responses at -80 mV in *stg/stg* cells (asterisk, $P < 0.05$). Responses at $+60$ mV are unaltered

($P > 0.1$). **e**, In *+stg* granule cells, stargazin Δ C transfection does not significantly alter the amplitude of glutamate-evoked responses at either -80 mV ($P > 0.1$) or $+60$ mV ($P > 0.1$). **f**, Transfection of stargazin Δ C into *+stg* granule cells downregulates synaptic AMPAR responses ($P < 1 \times 10^{-7}$). Cells were recorded in pairs in an attempt to control for differences in connectivity that may occur across cover slips and culture batches. WT, wild type.

labelling, but no decrease in NR1 labelling of cerebellar mossy fibre granule cell synapses in *stg/stg* mice. Expressing stargazin in *stg/stg* granule cells rescues AMPAR responses to synaptically released glutamate. Finally, when transfected into cultured hippocampal neurons, stargazin–GFP was found to cluster at synapses and colocalize with GluR1.

Further studies indicated that the PDZ-binding domain of stargazin is required for a subset of its actions. Removing the C-terminal PDZ binding domain from stargazin disrupts binding to PSD-95 and presumably other PDZ proteins in heterologous cells, but has no effect on AMPAR binding. This stargazin Δ C mutant rescues AMPAR responses to exogenous glutamate in *stg/stg* granule cells, but fails to restore synaptic responses. Finally, stargazin Δ C does not cluster at synapses and severely impairs AMPAR synaptic responses in hippocampal neurons.

These findings suggest that stargazin has two distinct roles in controlling AMPAR function. First, stargazin regulates delivery of AMPARs to the membrane surface and this function does not require the PDZ-binding domain. Second, stargazin mediates synaptic targeting of AMPARs and this function does require the PDZ-binding C terminus. Moreover, we propose that stargazin and the related proteins γ -3 and γ -4 mediate AMPAR targeting throughout the brain^{12,13}. That stargazin clusters at hippocampal synapses and that stargazin Δ C acts as a dominant-negative support a general role for stargazin-like proteins in synaptic targeting of

AMPARs. A class of PDZ-domain-containing proteins, including GRIP (glutamate receptor interacting protein) 1 and 2, ABP (AMPA-binding protein), PICK1 (protein interacting with C kinase 1) and SAP-97 (synapse-associated protein) have been implicated in synaptic targeting/clustering of AMPARs^{19–21}. A recent study suggests that the association of GluR2 with ABP/GRIP is not essential for synaptic targeting, but is required for maintaining the synaptic surface accumulation of AMPARs²². Together, it seems that synaptic targeting/insertion and synaptic stabilization of AMPARs may be mediated by several mechanisms. Whereas the interaction between stargazin and PSD-95 or related PDZ proteins is crucial for the initial synaptic targeting of AMPARs, a different pathway involving the association of GluR2 with ABP/GRIP may be required for receptor stabilization.

A number of issues remain unanswered. The site(s) on the AMPARs and stargazin responsible for binding has not been identified and may involve the transmembrane domains. It is unlikely that the interaction between stargazin and AMPARs occurs within the intracellular C-terminal domain of AMPARs, as very little homology of this domain exists among GluR1, -2 and -4. How stargazin controls surface expression of AMPARs in granule cells is unknown. Stargazin might act as a chaperone or might mask a retention signal on the AMPARs. However, because surface expression of AMPARs can be detected in many cells including oocytes, HEK293 cells and so on, it would seem that either all these

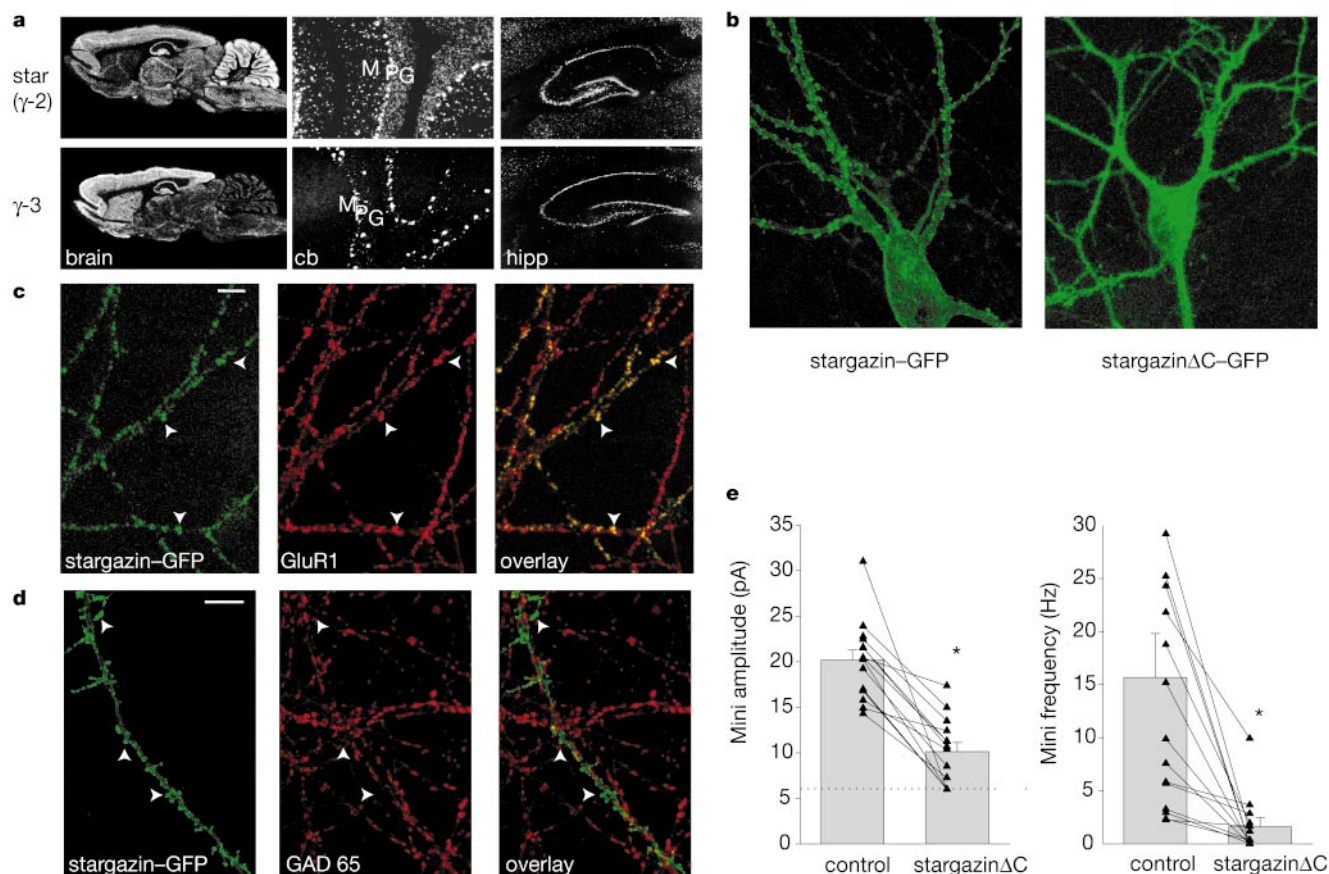


Figure 7 Stargazin Δ C downregulates hippocampal excitatory synapses. **a**, *In situ* hybridization shows discrete neuronal localizations of stargazin (γ -2) and γ -3 in brain. In cerebellum (cb) stargazin occurs in granule cells (G), in Purkinje neurons (P) and in stellate cells of the molecular layer (M). By contrast, γ -3 occurs in Golgi cells of granule cell layer and in some Purkinje neurons, but is absent from granule cells and molecular layer. In hippocampus (hipp), stargazin and γ -3 are co-expressed in similar neuronal populations. **b**, Stargazin–GFP, when transfected into hippocampal neurons, exhibits a punctate

distribution, whereas stargazin Δ C–GFP shows a diffuse pattern. Scale bar, 10 μ m. **c**, Stargazin–GFP colocalizes with GluR1 in 12-DIV cultures. Scale bar, 10 μ m. **d**, GAD 65 staining does not overlap with stargazin–GFP. Scale bar, 10 μ m. **e**, Stargazin Δ C transfection downregulates AMPAR mEPSC amplitude ($P < 1 \times 10^{-6}$) and frequency ($P < 0.05$) in hippocampal neurons. For the analysis of mEPSC amplitude, a 6-pA detection threshold (dashed line) was used: cells with no detectable mEPSCs were assigned a value of 6 pA.

cell types possess a stargazin-like protein or that high GluR expression can bypass this regulation. Finally, although NMDAR function is independent of stargazin, kainate receptors, which are structurally similar to AMPARs, might require stargazin for normal trafficking.

We have identified an essential role for stargazin in the synaptic targeting of AMPARs. We raise the intriguing possibility that stargazin-like proteins may participate in rapid activity-dependent trafficking of AMPARs in synaptic plasticity, a proposal supported by the known impairment of stargazer mouse in motor learning²³. □

Methods

Antibodies

The following antibodies were used: rabbit polyclonal antibodies to PSD-95, SAP-97, Kv1.4 (ref. 17), GluR1, GluR2 and GluR4 (Chemicon); monoclonal antibodies to PSD-95 (Affinity Bioreagents), NR1 (Pharmingen), GFP (Quantum, Clontech), haemagglutinin A (HA) epitope tag (BABC0) and Myc epitope tag (Santa Cruz); mouse monoclonal antibodies to synaptophysin (Sigma) and GAD 65 (a gift from S. Baeskoskov). The antibody to stargazin was generated by injecting rabbits with a fusion protein of the stargazin C-terminal 120 amino acids. The primary antibodies for the electron microscopy study (GluR2/3, GluR4, NR1 polyclonal and monoclonal) have been characterized^{24–27}.

Plasmid constructs

Stargazin cDNA (generously provided by V. Letts) was PCR amplified and subcloned into eukaryotic expression vectors pCDNA3 (Stratagene) and GW1-CMV (British Biotechnology). The HA epitope was inserted into Stargazin at the unique *Xho*I site (between residues 105 and 106). EGFP was inserted into Stargazin at the *Bgl*II site (between residues 269–270). GluR4 was tagged with GFP by inserting GFP between amino acids 3 and 4 of the mature protein.

Tissue culture and immunolabelling

Transfection (Lipofectamine Plus) and immunolabelling of COS cells were done as described²⁸. For analysis of surface GluR, COS-7 cells transfected with GluR4–GFP were washed in ice-cold DMEM and then incubated with DMEM containing anti-GFP antibody (Clontech) for 30 min at 4 °C before fixation in 2% paraformaldehyde. For immunoprecipitation, transfected COS cells were washed with ice-cold PBS and resuspended in 0.5 ml of lysis buffer containing TEE (50 mM Tris-HCl pH 7.4, 1 mM EDTA, 1 mM EGTA), 150 mM NaCl, 1% Triton X100, PMSF, aprotinin and leupeptin. Cells were solubilized by brief sonication followed by extraction for 30 min at 4 °C. Insoluble material was removed by centrifugation at 10,000g for 10 min. Samples were then incubated with GFP antibodies for 1 h at 4 °C. After addition of 20 µl of Protein A-Sepharose beads (Pharmacia), samples were incubated for 1 h at 4 °C.

Immunoprecipitates were washed three times with buffer containing TEE, 150 mM NaCl and 1% Triton X-100, boiled in SDS–PAGE sample buffer with 1 mM dithiothreitol for 2 min and analysed by SDS–PAGE. Input was 5% of the co-immunoprecipitation. Hippocampal primary cultures were prepared from newborn Sprague–Dawley rats as described²⁹, except that B27-supplement (Life Technologies) was added to the culture media. Stargazer breeding pairs were from Jackson Laboratory. Stargazer cerebellar granule cell cultures were prepared from 5–7-day-old mouse pups as described¹⁴. Cells were in minimum essential medium (5.3 mM K⁺; Gibco), supplemented with glucose, transferrin, insulin, glutamine, cytosine arabinoside and 10% heat-inactivated fetal calf serum (Gibco). Tail samples from individual pups were used for genotyping with described primers¹¹. Neuronal cultures were transfected using DOTAP (Roche) or Effectene (Qiagen). Immunocytochemistry was performed on 8–9-DIV granule cell cultures or 12–14-DIV hippocampal cultures. After fixation with methanol (15 min, –20 °C) or 2% paraformaldehyde (15 min, room temperature), cells were washed with PBS containing 0.3% Triton X100 (PBST) before being incubated with primary antibody and secondary antibodies. Imaging used either epifluorescence with a ×63 oil-immersion objective or confocal microscopy with a ×60 oil-immersion objective.

Subcellular fractionation

Subcellular fractions of rat brains were prepared by differential centrifugation as described³⁰.

Electrophysiology

Whole-cell patch-clamp recordings were made at room temperature using 3–7-MΩ patch pipettes filled with an internal solution containing (in mM) 140 CsCl, 2 MgCl₂, 5 EGTA, 10 HEPES, 0.3 Na₃-GTP, 4 Na₂-ATP pH 7.35. Cultures were continuously superfused with external solution containing (in mM) 119 NaCl, 26 NaHCO₃, 2.5 KCl, 10 glucose, 2.5 CaCl₂, 1.3 MgSO₄, 1 NaH₂PO₄, 0.1 picrotoxin. Tetrodotoxin (TTX; 1 µM) was added for agonist application experiments and miniature EPSC recordings. During spontaneous and miniature EPSC recordings, cells were held at –60 mV (unless otherwise stated). For most experiments, neighbouring transfected and untransfected cells were selected for recording in an attempt to minimize any differences that may occur across cover slips and platings, such as the degree of connectivity. Fast agonist application was achieved by gravity feeding as described³¹. Calcium current recordings were obtained from 2-DIV granule cell cultures. External solution contains (in mM) 110 NaCl, 20 TEA-Cl, 26 NaHCO₃, 2.5 KCl, 10 glucose, 3.0 CaCl₂, 1.3 MgSO₄, 1 NaH₂PO₄, 0.1 picrotoxin and 1 µM

TTX. The patch-pipette solution contains (in mM) 130 CsCl, 20 TEA-Cl, 2 MgCl₂, 5 EGTA, 10 HEPES, 0.3 Na₃-GTP, 4 Na₂-ATP. Current records were filtered at 2 kHz and digitized at 5 kHz.

Postembedding immunogold labelling

The postembedding immunogold method has been described^{32–35} and is modified from an earlier method³⁶. Parasagittal sections of the dorsostral quadrant of the cerebellum were cryoprotected and frozen in a Leica EM CPS, then placed in a Leica AFS, infiltrated with Lowicryl HM 20, and polymerized with ultraviolet light. Thin sections were immunolabelled using 10-nm immunogold (Amersham). Glomerular zones of the granular layer were selected and photographed from a wide area of the section. All synapses in the micrographs were counted. For both *stg/stg* and *+stg*, most of these synapses showed ultrastructural features typical of granule cell dendrite synapses³⁷. All immunogold particles from the postsynaptic density and cleft were counted, as described^{32–35}.

Received 17 July; accepted 31 October 2000.

- Luscher, C., Nicoll, R. A., Malenka, R. C. & Muller, D. Synaptic plasticity and dynamic modulation of the postsynaptic membrane. *Nature Neurosci.* **3**, 545–550 (2000).
- Malenka, R. C. & Nicoll, R. A. Long-term potentiation—a decade of progress? *Science* **285**, 1870–1874 (1999).
- Malinow, R., Mainen, Z. F. & Hayashi, Y. LTP mechanisms: from silence to four-lane traffic. *Curr. Opin. Neurobiol.* **10**, 352–357.
- Liu, S. Q. & Cull-Candy, S. G. Synaptic activity at calcium-permeable AMPA receptors induces a switch in receptor subtype. *Nature* **405**, 454–458 (2000).
- Luthi, A. *et al.* Hippocampal LTD expression involves a pool of AMPARs regulated by the NSF–GluR2 interaction. *Neuron* **24**, 389–399 (1999).
- Man, Y. H. *et al.* Regulation of AMPA receptor-mediated synaptic transmission by clathrin-dependent receptor internalization. *Neuron* **25**, 649–662 (2000).
- Song, I. *et al.* Interaction of the N-ethylmaleimide-sensitive factor with AMPA receptors. *Neuron* **21**, 393–400 (1998).
- P. Osten, S. *et al.* The AMPA Receptor GluR2 C terminus can mediate a reversible, ATP-dependent interaction with NSF and SNAPs. *Neuron* **21**, 99–110 (1998).
- Chen, L., Bao, S., Qiao, X. & Thompson, R. F. Impaired cerebellar synapse maturation in *waggle*, a mutant mouse with a disrupted neuronal calcium channel gamma subunit. *Proc. Natl Acad. Sci. USA* **96**, 12132–12137 (1999).
- Hashimoto, K. *et al.* Impairment of AMPA receptor function in cerebellar granule cells of ataxic mutant mouse stargazer. *J. Neurosci.* **19**, 6027–6036 (1999).
- Letts, V. A. *et al.* The mouse stargazer gene encodes a neuronal Ca²⁺-channel gamma subunit. *Nature Genet.* **19**, 340–347 (1998).
- Burgess, D. L., Davis, C. F., Gefrides, L. A. & Noebels, J. L. Identification of three novel Ca²⁺ channel gamma subunit genes reveals molecular diversification by tandem and chromosome duplication. *Genome Res.* **9**, 1204–1213 (1999).
- Klugbauer, N. *et al.* A family of gamma-like calcium channel subunits. *FEBS Lett.* **470**, 189–197 (2000).
- Randall, A. & Tsien, R. W. Pharmacological dissection of multiple types of Ca²⁺ channel currents in rat cerebellar granule neurons. *J. Neurosci.* **15**, 2995–3012 (1995).
- Hsueh, Y. P. *et al.* Direct interaction of CASK/LIN-2 and syndecan heparan sulfate proteoglycan and their overlapping distribution in neuronal synapses. *J. Cell Biol.* **142**, 139–151 (1998).
- Torres, R. *et al.* PDZ proteins bind, cluster, and synaptically colocalize with Eph receptors and their ephrin ligands. *Neuron* **21**, 1453–1463 (1998).
- Kim, E., Niethammer, M., Rothschild, A., Jan, Y. N. & Sheng, M. Clustering of Shaker-type K⁺ channels by interaction with a family of membrane-associated guanylate kinases. *Nature* **378**, 85–88 (1995).
- Monyer, H., Burnashev, N., Laurie, D. J., Sakmann, B. & Seeburg, P. H. Developmental and regional expression in the rat brain and functional properties of four NMDA receptors. *Neuron* **12**, 529–540 (1994).
- Sheng, M. & Pak, D. T. Glutamate receptor anchoring proteins and the molecular organization of excitatory synapses. *Ann. N. Y. Acad. Sci.* **868**, 483–493 (1999).
- Garner, C. C., Nash, J. & Huganir, R. L. PDZ domains in synapse assembly and signalling. *Trends Cell Biol.* **10**, 274–280 (2000).
- Braithwaite, S. P., Meyer, G. & Henley, J. M. Interactions between AMPA receptors and intracellular proteins. *Neuropharmacology* **39**, 919–930 (2000).
- Osten, P. *et al.* Mutagenesis reveals a role for ABP/GRIP binding to GluR2 in synaptic surface accumulation of the AMPA receptor. *Neuron* **27**, 313–325 (2000).
- Bao, S., Chen, L., Qiao, X., Knusel, B. & Thompson, R. F. Impaired eye-blink conditioning in *waggle*, a mutant mouse with cerebellar BDNF deficiency. *Learn. Mem.* **5**, 355–364 (1998).
- Wentholt, R. J., Yokotani, N., Doi, K. & Wada, K. Immunochemical characterization of the non-NMDA glutamate receptor using subunit-specific antibodies. Evidence for a hetero-oligomeric structure in rat brain. *J. Biol. Chem.* **267**, 501–507 (1992).
- Petralia, R. S. & Wentholt, R. J. Light and electron immunocytochemical localization of AMPA-selective glutamate receptors in the rat brain. *J. Comp. Neurol.* **318**, 329–354 (1992).
- Petralia, R. S., Yokotani, N. & Wentholt, R. J. Light and electron microscope distribution of the NMDA receptor subunit NMDAR1 in the rat nervous system using a selective anti-peptide antibody. *J. Neurosci.* **14**, 667–696 (1994).
- Luo, J., Wang, Y., Yasuda, R. P., Dunah, A. W. & Wolfe, B. B. The majority of N-methyl-D-aspartate receptor complexes in adult rat cerebral cortex contain at least three different subunits (NR1/NR2A/NR2B). *Mol. Pharmacol.* **51**, 79–86 (1997).
- El-Husseini, A. E. *et al.* Dual palmitoylation of PSD-95 mediates its vesiculotubular sorting, postsynaptic targeting, and ion channel clustering. *J. Cell Biol.* **148**, 159–172 (2000).
- Lester, R. A. J., Quarum, M. L., Parker, J. D., Weber, E. & Jahr, C. E. Interaction of 6-cyano-7-nitroquinoxaline-2, 3-dione with the N-methyl-D-aspartate receptor-associated glycine binding site. *Mol. Pharmacol.* **35**, 565–570 (1989).
- Jo, K., Derin, R., Li, M. & Bredt, D. S. Characterization of MALS/Vel1-1, -2, and -3: a family of mammalian LIN-7 homologs enriched at brain synapses in association with the postsynaptic density-95/NMDA receptor postsynaptic complex. *J. Neurosci.* **19**, 4189–4199 (1999).

31. Lester, R. A. J. & Jahr, C. E. NMDA channel behavior depends on agonist affinity. *J. Neurosci.* **12**, 635–643 (1992).
32. Wang, Y. -X., Wenthold, R. J., Ottersen, O. P. & Petralia, R. S. Endbulb synapses in the anteroventral cochlear nucleus express a specific subset of AMPA-type glutamate receptor subunits. *J. Neurosci.* **18**, 1148–1160 (1998).
33. Petralia, R. S., Zhao, H. M., Wang, Y. X. & Wenthold, R. J. Variations in the tangential distribution of postsynaptic glutamate receptors in Purkinje cell parallel and climbing fibre synapses during development. *Neuropharmacology* **37**, 1321–1334 (1998).
34. Petralia, R. S. *et al.* Selective acquisition of AMPA receptors over postnatal development suggests a molecular basis for silent synapses. *Nature Neurosci.* **2**, 31–36 (1999).
35. Zhao, H. M., Wenthold, R. J. & Petralia, R. S. Glutamate receptor targeting to synaptic populations on Purkinje cells is developmentally regulated. *J. Neurosci.* **18**, 5517–5528 (1998).
36. Matsubara, A., Laake, J. H., Davanger, S., Usami, S. & Ottersen, O. P. Organization of AMPA receptor subunits at a glutamate synapse: a quantitative immunogold analysis of hair cell synapses in the rat organ of Corti. *J. Neuroscience* **16**, 4457–4467 (1996).
37. Palay, S. L. & Chan-Palay, V. *Cerebellar Cortex. Cytology and Organization* (Springer, New York, 1974).

Acknowledgements

We thank R. F. Thompson and X. Qiao for providing the initial stargazer breeding pairs; K. P. Campbell for providing a stargazin antibody used in preliminary experiments; B. B. Wolfe for providing the NR1 monoclonal antibody used in the electron microscopy study; S. Tomita for subcloning Stargazin-3; E. Schnell for assisting in hippocampal culture transfection; Q. Zhou for assisting in confocal microscopy; and Y.-X. Wang for assisting in immunogold labelling. We also thank M. Frerking, L. Jan and D. Julius for their comments on the manuscript. D.S.B is supported by grants from NIH and HHMI. R.A.N. is supported by grants from the NIH and Bristol-Myers Squibb. D.S.B. is an established investigator of the American Heart Association, D.M.C. is a postdoctoral fellow of the HHMI. R.A.N. is a member of the Keck Center for Integrative Neuroscience and the Silvio Conte Center for Neuroscience Research.

Correspondence and requests for materials should be addressed to R.A.N. (e-mail: nicoll@phy.ucsf.edu).



Electric-field control of ferromagnetism

H. Ohno, D. Chiba, F. Matsukura, T. Omiya, E. Abe, T. Dietl*, Y. Ohno & K. Ohtani

Laboratory for Electronic Intelligent Systems, Research Institute of Electrical Communication, Tohoku University, Katahira 2-1-1, Aoba-ku, Sendai 980-8577, Japan

* Present address: Institute of Physics and College of Science, Polish Academy of Sciences, al. Lotników 32/46, PL-02668 Warszawa, Poland

It is often assumed that it is not possible to alter the properties of magnetic materials once they have been prepared and put into use. For example, although magnetic materials are used in information technology to store trillions of bits (in the form of magnetization directions established by applying external magnetic fields), the properties of the magnetic medium itself remain unchanged on magnetization reversal. The ability to externally control the properties of magnetic materials would be highly desirable from fundamental and technological viewpoints, particularly in view of recent developments in magnetoelectronics and spintronics^{1,2}. In semiconductors, the conductivity can be varied by applying an electric field, but the electrical manipulation of magnetism has proved elusive. Here we demonstrate electric-field control of ferromagnetism in a thin-film semiconducting alloy, using an insulating-gate field-effect transistor structure. By applying electric fields, we are able to vary isothermally and reversibly the transition temperature of hole-induced ferromagnetism.

(In,Mn)As, used as the magnetic channel material, is one of the ferromagnetic III–V semiconductors^{3–6} and is chosen for the following two reasons: (1) it exhibits hole-induced ferromagnetism (transition temperature $T_C \approx 30$ K or below)^{3,4} and photo-carrier-induced magnetic transitions⁷ at relatively low carrier concentration, which is important for the gate control of carrier concentration; and (2) it can be epitaxially grown within non-magnetic semiconducting InAs/(Al,Ga)Sb heterostructures, thereby allowing precise control of nonmagnetic–magnetic structures. Manganese substitutes indium in (In,Mn)As^{8,9} and simultaneously provides a localized magnetic moment and a hole, owing to its acceptor nature. These holes mediate magnetic interaction, resulting in ferromagnetism¹⁰. Because a weak antiferromagnetism caused by superexchange is observed in the absence of holes¹¹, the decrease (increase) of the hole concentration by the application of electric

fields is expected to result in a reduction (an increase) of the hole-mediated ferromagnetic exchange interaction among localized Mn spins, and hence in the modification of ferromagnetic properties as shown in Fig. 1, in the cross-section of the metal-insulator–semiconductor field-effect transistor (FET) structure studied here.

The semiconductor structure consists of, from the surface side, 5 nm of (In,Mn)As (with Mn composition 0.03), 10 nm of InAs, 400 nm of (Al,Ga)Sb (with Al composition 0.6) and 100 nm of AlSb on semi-insulating (001) GaAs substrates. The AlSb and (Al,Ga)Sb layers serve as a buffer structure, in which most of the 7% lattice-mismatch between the GaAs substrate and the epitaxial layers is relaxed. The thin InAs buffer layer ensures the two-dimensional growth of the subsequent (In,Mn)As layer and improves its magnetic quality¹². All the layers are grown by solid-source molecular beam epitaxy; the AlSb and (Al,Ga)Sb layers are grown at 550 °C, the InAs layer at 450 °C, and the magnetic layer at 250 °C. A Hall bar geometry mesa is first formed by Cl dry-etch removal of the unwanted epitaxial layers. Then a 0.8- μ m thick polyimide layer (relative dielectric constant 3.3) is spun-on, which acts as a gate insulator, and contact holes for source and drain are formed. Finally, the Cr(5 nm)/Au(95 nm) gate electrode is defined by a lift-off process and the source and drain contacts are made by indium bonding. The gate leakage currents are less than 5 nA cm^{–2} in all the devices. The potential probes of the Hall bar are separated by 80 μ m and the width of the channel is 60 μ m. Below, we show typical results obtained from over 40 FETs fabricated thus far.

The small dimension of the device together with the thick diamagnetic GaAs substrate makes it difficult to measure directly the magnetization of the magnetic layer. We use the magnetization-dependent Hall effect to extract the magnetic properties of the channel layer. The sheet Hall resistivity, R_{Hall} , is given by the sum of the ordinary Hall effect due to the Lorentz force and the anomalous Hall effect originating from the asymmetric scattering in the presence of magnetization:

$$R_{\text{Hall}} = \frac{R_0}{d} B + \frac{R_s}{d} M$$

where R_0 is the ordinary Hall coefficient, B and M the magnetic field and the magnetization perpendicular to the layer, respectively, R_s the anomalous Hall coefficient, and d the thickness of the channel

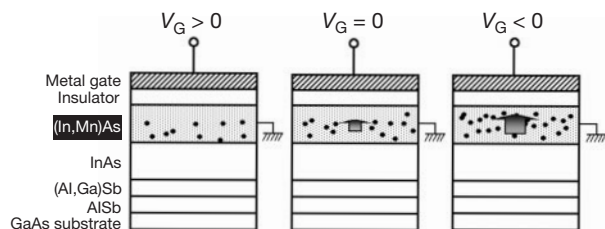


Figure 1 Field-effect control of the hole-induced ferromagnetism in magnetic semiconductor (In,Mn)As field-effect transistors. Shown are the cross-sections of a metal-insulator–semiconductor structure under gate biases V_G . This controls the hole concentration in the magnetic semiconductor channel (filled circles). Negative V_G increases hole concentration, resulting in enhancement of the ferromagnetic interaction among magnetic Mn ions, whereas positive V_G has an opposite effect. The arrow schematically shows the magnitude of the Mn magnetization. The InAs/(Al,Ga)Sb/AlSb structure under the (In,Mn)As layer serves as a buffer relaxing the lattice mismatch between the structure and the GaAs substrate to produce a smooth surface on which the magnetic layer is grown.

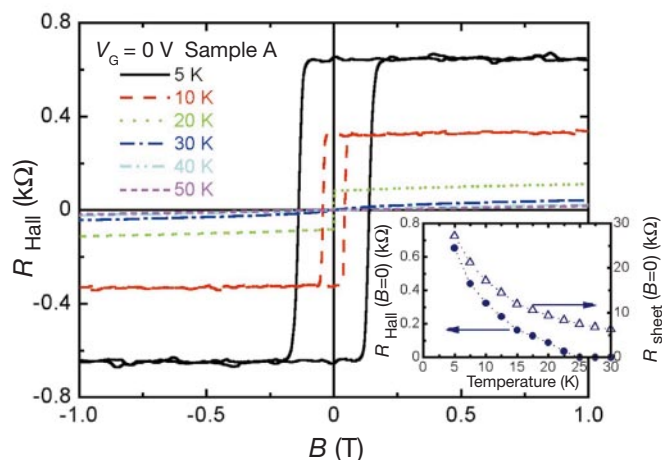


Figure 2 Magnetic-field dependence of the sheet Hall resistance R_{Hall} proportional to the magnetization of the magnetic semiconductor layer. R_{Hall} is used to measure the small magnetization of the channel. Shown are R_{Hall} as a function of field perpendicular to the layer at temperatures $T = 5$ –50 K of sample A at $V_G = 0$ V. Clear hysteresis observed at $T \lesssim 20$ K is evidence of ferromagnetism. Inset, the temperature dependence of the remanence of R_{Hall} (solid circles), showing that the ferromagnetic transition temperature T_C is above 20 K. Open circles indicate the channel sheet resistance R_{sheet} at zero field, which shows moderate negative T -dependence.

layer¹³. R_S is proportional to ρ^γ , where ρ is the resistivity of (In,Mn)As. Normally γ is either 1 or 2 (ref. 13). The R_{Hall} versus B curves of sample A with the gate bias $V_G = 0$ V are shown in Fig. 2. The clear square hysteresis at low temperatures $T \lesssim 20$ K shows that the (In,Mn)As layer is ferromagnetic with perpendicular easy axis⁴ and $R_{\text{Hall}} \propto M$, enabling us to extract the magnetization of the channel layer. At room temperature, the ordinary Hall effect dominates because of the much reduced susceptibility of the channel. Temperature dependence of the remanence in R_{Hall} (R_{Hall} at $B = 0$ T) shown in the inset indicates that T_C is above 20 K (T_C varies by about 20% from device to device). The inset also shows that the channel sheet resistance R_{sheet} at $B = 0$ T gradually increases as temperature decreases. In the following experiments, we apply $V_G = \pm 125$ V, which modulates ± 4 –6% of R_{sheet} in the range of $T = 5$ –300 K. From the gate capacitance, $|V_G| = 125$ V is expected to produce a sheet hole concentration change of $3 \times 10^{12} \text{ cm}^{-2}$; hence, assuming a constant mobility, the total sheet hole concentration is 5 – $8 \times 10^{13} \text{ cm}^{-2}$, in agreement with the Hall concentration at 300 K.

The R_{Hall} versus B near T_C at $T = 22.5$ K of sample B under three different V_G measured in the sequence of $V_G = 0, +125, -125$ and 0 V are displayed in Fig. 3. The measurements are done with a small constant current flow, where no current dependence is present. At $V_G = 0$ V, a moderate hysteresis is observed in the curve, confirming that the sample temperature is close to but below T_C . Application of $V_G = +125$ V (in the direction of depletion of holes) transforms the moderate hysteresis loop to a paramagnetic response with reduced susceptibility and no hysteresis. Switching V_G to -125 V changes the response to the one with a clear square hysteresis. The original R_{Hall} – B curve is restored when V_G is returned to 0 V. The inset shows the same curves to higher magnetic fields. The qualitative difference of the magnetization curves obtained in Fig. 3 indicates that the magnetism of the channel layer is modified isothermally in a reversible way by the external gate voltage.

Figure 4 shows the spontaneous R_{Hall} (R_{Hall}^S), which is proportional to the spontaneous magnetization M_S , around T_C of sample A under the three gate biases ($V_G = +125, 0$, and -125 V). Because T_C is the temperature at which M_S (and hence R_{Hall}^S) becomes zero, Fig. 4

clearly shows that the applied gate bias changes T_C by ± 1 K ($\Delta T_C/T_C = \pm 4\%$), and in this temperature range, ferromagnetism can be turned on and off by the external electric fields. R_{Hall}^S , shown in Fig. 4, is determined from the extrapolation of R_{Hall} from the moderate fields ($0.1 < B < 0.7$ T) to $B = 0$ using an Arrott plot to minimize the effects of domain wall formation and magnetic anisotropy¹⁴. The inset of Fig. 4 shows the typical Arrott plots of sample A near T_C , where the intercept of a linear extrapolation of R_{Hall}^2 at $B = 0$ gives $(R_{\text{Hall}}^S)^2$. A negative intercept means no M_S . The narrow range of relevant temperatures and fields allows us to neglect the small T and B dependences of ρ , making the use of R_{Hall} instead of M_S an excellent approximation. The linear dependence seen in the inset shows that the contribution from the ordinary Hall effect is negligible.

Because the magnetic layer is very thin, it acts as a quantum well and V_G is believed to modulate the hole population of the two-dimensional (2D) sub-bands in the magnetic layer, as opposed to the depletion-layer movement in the three-dimensional (3D) FETs. Recent mean-field theory predicts no carrier concentration dependence of T_C for 2D systems, when only the ground sub-band with constant effective mass is occupied, because the 2D density-of-states that determines T_C is energy-independent¹⁵. As demonstrated in the theory for the 3D case¹⁰, however, the realistic valence band structure including spin-orbit coupling is important to determine T_C . Because the valence band density-of-states is a complex function of energy, we expect modulation of T_C through modulation of carrier concentration even in the 2D systems. (From the 3D theory¹⁰, we expect $\Delta T_C/T_C = 10\%$ for $\Delta p/p = 0.06$, in reasonable agreement with the present results). Additional contributions to the modulation of T_C may come from the particular design of the heterostructure employed here: a non-magnetic InAs is placed beneath the magnetic (In,Mn)As layer. Application of positive (negative) bias distorts the relevant 2D wavefunctions away from (towards) the magnetic layer, resulting in reduction (increase) of carrier–magnetic-impurity exchange interaction and hence in reduction (increase) of T_C as discussed in ref. 16.

Field-controlled ferromagnetism is expected to have a great technological impact because of the possible integration of ferromagnetic devices with non-magnetic III–V devices such as lasers

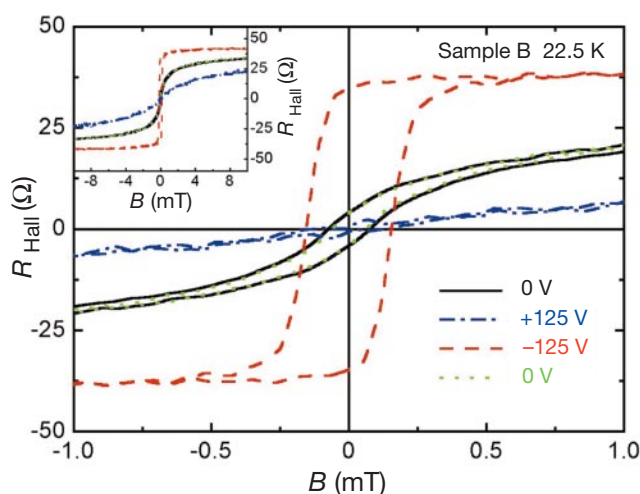


Figure 3 R_{Hall} versus field curves under three different gate biases. Application of $V_G = 0, +125$ and -125 V results in qualitatively different field dependence of R_{Hall} measured at 22.5 K (sample B). When holes are partially depleted from the channel ($V_G = +125$ V), a paramagnetic response is observed (blue dash-dotted line), whereas a clear hysteresis at low fields (< 0.7 mT) appears as holes are accumulated in the channel ($V_G = -125$ V, red dashed line). Two R_{Hall} curves measured at $V_G = 0$ V before and after application of ± 125 V (black solid line and green dotted line, respectively) are virtually identical. Inset, the same curves shown at higher magnetic fields.

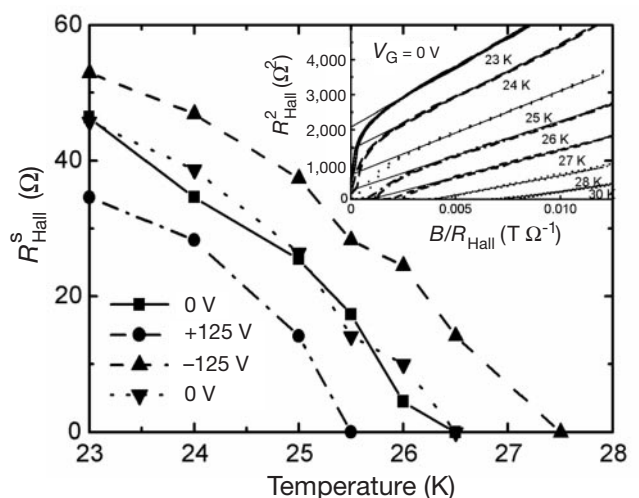


Figure 4 Temperature dependence of spontaneous Hall resistance R_{Hall}^S under three different gate biases. R_{Hall}^S proportional to the spontaneous magnetization M_S indicates ± 1 K modulation of T_C upon application of $V_G = \pm 125$ V (sample A). T_C is the temperature at which R_{Hall}^S (and hence M_S) goes to zero. Data at $V_G = 0$ V before and after application of ± 125 V are shown by squares and down triangles, respectively. In order to minimize the effect of domain rotation and magnetic anisotropy, R_{Hall}^S is determined by extrapolation of R_{Hall} from moderate fields (0.1 – 0.7 T) to 0 using Arrott plots (R_{Hall}^2 versus B/R_{Hall} plots shown in inset).

and transistors already in use in electronics. We note that room temperature ferromagnetism in semiconductors has been predicted¹⁰. Combinations of semiconductor spintronic devices^{17,18} with the present field-controlled ferromagnetism may also be important in quantum information technologies that are based on manipulation of spin states in semiconductors^{19,20}. □

Received 5 September; accepted 1 November 2000.

1. Prinz, G. A. Magnetoelectronics. *Science* **282**, 1660–1663 (1998).
2. de Boeck, J. & Borghs, G. Magnetoelectronics. *Phys. World* **12**, 27–32 (1999).
3. Ohno, H., Munekata, H., Penney, T., von Molnár, S. & Chang, L. L. Magnetotransport properties of p-type (In,Mn)As diluted magnetic III-V semiconductors. *Phys. Rev. Lett.* **68**, 2664–2667 (1992).
4. Munekata, H., Zaslavsky, A., Fumagalli, P. & Gambino, R. J. Preparation of (In,Mn)As/(Ga,Al)Sb magnetic semiconductor heterostructures and their ferromagnetic characteristics. *Appl. Phys. Lett.* **63**, 2929–2931 (1993).
5. Ohno, H. Making nonmagnetic semiconductors ferromagnetic. *Science* **281**, 951–956 (1998).
6. Ohno, H. Properties of ferromagnetic III-V semiconductors. *J. Mag. Magn. Mater.* **200**, 110–129 (1999).
7. Koshihara, S. *et al.* Ferromagnetic order induced by photogenerated carriers in magnetic III-V semiconductor heterostructures of (In,Mn)As/GaSb. *Phys. Rev. Lett.* **78**, 4617–4620 (1997).
8. Munekata, H. *et al.* Diluted magnetic III-V semiconductors. *Phys. Rev. Lett.* **63**, 1849–1852 (1989).
9. Soo, Y. L., Huang, S. W., Ming, Z. H., Kao, Y. H. & Munekata, H. III-V diluted magnetic semiconductor: Substitutional doping of Mn in InAs. *Phys. Rev. B* **53**, 4905–4909 (1996).
10. Dietl, T., Ohno, H., Matsukura, F., Cibert, J. & Ferrand, D. Zener model description of ferromagnetism in zinc-blende magnetic semiconductors. *Science* **287**, 1019–1022 (2000).
11. von Molnár, S., Munekata, H., Ohno, H. & Chang, L. L. New diluted magnetic semiconductor based on III-V compounds. *J. Mag. Magn. Mater.* **93**, 356–364 (1991).
12. Shen, A. *et al.* Epitaxy and properties of InMnAs/AlGaSb diluted magnetic III-V semiconductor heterostructures. *Appl. Surf. Sci.* **113/114**, 183–188 (1997).
13. Chien, C. L. & Westgate, C. W. *The Hall Effect and Its Applications* 43–51 (Plenum, New York, 1980).
14. Arrott, A. Criterion for ferromagnetism from observations of magnetic isotherms. *Phys. Rev.* **108**, 1394–1395 (1957).
15. Dietl, T., Haury, A. & Merle d'Aubigné, Y. Free carrier-induced ferromagnetism in structures of diluted magnetic semiconductors. *Phys. Rev. B* **55**, R3347–R3350 (1997).
16. Lee, B., Jungwirth, T. & MacDonald, A. H. Theory of ferromagnetism in diluted magnetic semiconductor quantum wells. *Phys. Rev. B* **61**, 15606–15609 (2000).
17. Fiederling, R. *et al.* Injection and detection of a spin-polarized current in a light-emitting diode. *Nature* **402**, 787–790 (1999).
18. Ohno, Y. *et al.* Electrical spin injection in a ferromagnetic semiconductor heterostructure. *Nature* **402**, 790–792 (1999).
19. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
20. Vrijen, R. *et al.* Electron-spin-resonance transistors for quantum computing in silicon-germanium heterostructures. *Phys. Rev. A* **62**, 012306–1–10 (2000).

Acknowledgements

This work was supported by the Japan Society for the Promotion of Science, the Ministry of Education, Japan, and the Mitsubishi Foundation.

Correspondence and requests for materials should be addressed to H.O. (e-mail: ohno@riec.tohoku.ac.jp).

Mesoscopic fast ion conduction in nanometre-scale planar heterostructures

N. Sata, K. Eberman*, K. Eberl & J. Maier

Max-Planck-Institut für Festkörperforschung, 70569 Stuttgart, Germany

Ion conduction is of prime importance for solid-state reactions in ionic systems, and for devices such as high-temperature batteries and fuel cells, chemical filters and sensors^{1,2}. Ionic conductivity in solid electrolytes can be improved by dissolving appropriate impurities into the structure or by introducing interfaces that cause the redistribution of ions in the space-charge regions^{3–11}. Heterojunctions in two-phase systems should be particularly

efficient at improving ionic conduction^{3,4}, and a qualitatively different conductivity behaviour is expected when interface spacing is comparable to or smaller than the width of the space-charge regions in comparatively large crystals^{12–15}. Here we report the preparation, by molecular-beam epitaxy, of defined hetero-layered films composed of CaF₂ and BaF₂ that exhibit ionic conductivity (parallel to the interfaces) increasing proportionally with interface density—for interfacial spacing greater than 50 nanometres. The results are in excellent agreement with semi-infinite space-charge calculations³, assuming a redistribution of fluoride ions at the interfaces. If the spacing is reduced further, the boundary zones overlap and the predicted mesoscopic size effect^{3,12} is observed. At this point, the single layers lose their individuality and an artificial ionically conducting material with anomalous transport properties is generated. Our results should lead to fundamental insight into ionic contact processes and to tailored ionic conductors of potential relevance for medium-temperature applications.

The growth of semiconductor heterostructures by molecular beam epitaxy (MBE) is the method of choice for preparing hetero-systems with highly defined geometry, periodicity, interfacial spacings and layer sequence^{16,17}. We used this technique, performed in a high-vacuum chamber with a base pressure of 10^{−9} mbar. Single crystals of the moderate (anti-Frenkel disordered) fluoride conductors CaF₂ and BaF₂ (refs 18–20) were sublimed at 1,180 and 1,000 °C, respectively, leading to a layer growth rate of about 1 nm min^{−1}. Periodicity and thickness were varied over a wide range (individual layer thicknesses: 2–500 nm) by computer-controlled effusion cell shutters. The total thickness of the film packages ranged from 200 to 500 nm. Single crystals of Al₂O₃ and SiO₂ were tested as substrate materials; they showed little influence on the conductivity of the thin films, whereas they obviously affected crystallinity. The substrate was kept at a temperature of 500 °C or about 100 °C to investigate the effect of growth temperature on film structure, which could also affect the ionic conductivity.

The films were characterized by X-ray diffraction and secondary-ion mass spectrometry (SIMS). In contrast to CaF₂, the BaF₂ film grows on Al₂O₃(012) substrates at a temperature of 500 °C in [111]

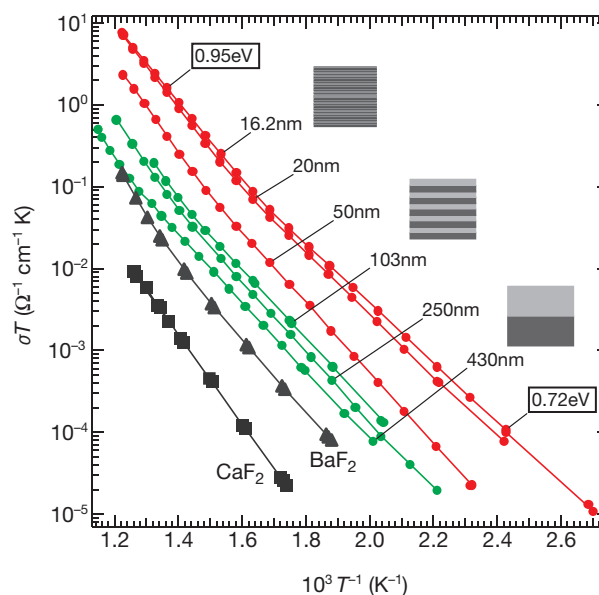


Figure 1 Parallel ionic conductivity of the films. Data are shown for films with various periods and interfacial densities in the 430–16 nm range. We note that the overall thickness is approximately the same in all cases (~500 nm). σ , conductivity; T , temperature. The different colours refer to different site regimes (green: semi-infinite space charge zones; red: finite space charge zones).

* Present address: 3M Center, St Paul, Minnesota 55144-1000, USA.

directions exclusively. Therefore, BaF_2 was deposited as the first layer. Such heterolayered films exhibited a uniform crystal orientation when prepared at high substrate temperatures (that is, about 500°C), as confirmed by X-ray pole-figure measurements, and by transmission electron microscopy (TEM). Films obtained at low substrate temperatures (about 100°C), or obtained from simultaneous BaF_2 and CaF_2 beams, are polycrystalline.

The SIMS spectra of the $\text{BaF}_2/\text{CaF}_2$ heterofilms showed clearly the layered character of the structure with minor mutual solubility. Although we detected oxygen and carbon impurities by SIMS, the fact that their amount did not vary strongly between the different samples excludes any major relevance to the effects considered. While TEM shows rather sharp boundaries, a certain degree of intermixing of Ba and Ca at the interface cannot be ruled out by the SIMS measurements, owing to the limited resolution.

Platinum films were deposited on the surface of the heterostructure as rectangular-shaped electrodes for lateral conductivity measurements. The 13-mm-long surface electrodes were separated by 1 mm, which may be compared with the film thicknesses of less than 10^{-3} mm. This geometry ensures that the parallel conductance of the total film was measured. Ionic conductivities were confirmed by impedance spectra, obtained by a.c. impedance measurements from 1 to 10^4 Hz at different temperatures over the range 540 – 100°C .

The results for the pure CaF_2 and BaF_2 films (black symbols in Fig. 1) show the intrinsic to extrinsic transition, as in the bulk, with the expected activation energies. But owing to the small sample thickness (~ 500 nm), the low-temperature conductivity need not, at least not exclusively, be due to impurity doping (for example, O^{2-} on F sites), as the space-charge contribution of the surface and the substrate interface would result in similar slopes^{9,12,21}. (We note the similarity of the BaF_2 data to the results of calculations involving two space-charge layers, which will be discussed below.)

Figures 1 and 2 reveal that large conductivities can be achieved in

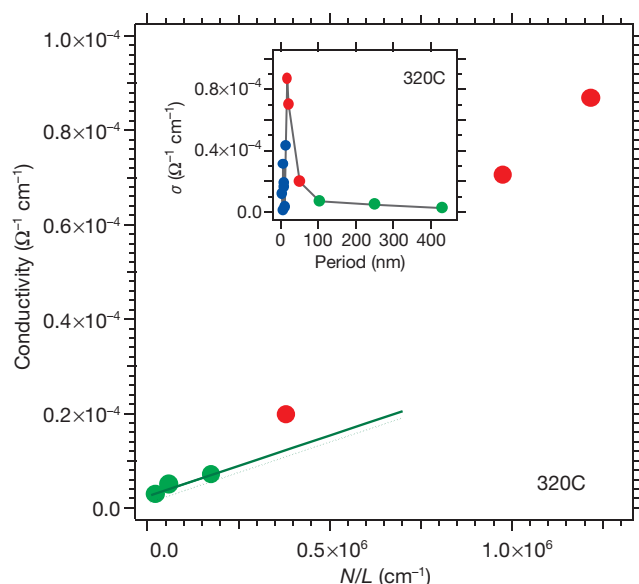


Figure 2 Variation of ionic conductivity with N/L . In the period range 430 – 100 nm (green), the (parallel) ionic conductivity increases monotonically with the number of interfaces (N). The linear increase with N/L (where L is overall thickness) exhibits exactly the slope obtained from the space-charge calculations. The prediction without any adjustable parameter (that is, ignoring background effects; dotted line) is very close to the absolute values. The bold line refers to the background-corrected calculation (red). The inset gives the conductivities as a function of the period in the range 430 – 2 nm. The samples with smaller spacings (blue) (~ 50 nm $\geq$ period ≥ 16 nm) show a further anomalous increase attributable to overlapping space charges. When the period is less than 16 nm (see inset), the conductivity decreases as discussed in the text.

the heterostructures. The examples given there are ... $\text{CaF}_2/\text{BaF}_2$... heterolayers composed of CaF_2 and BaF_2 units of uniform and equal thicknesses, with varying interfacial densities while the overall film thickness L was always similar (470 ± 40 nm).

All experimental results shown in Fig. 1 are stationary values observed after annealing at 540°C . The first cycles suffer from non-equilibrium effects possibly due to non-stationary dislocations. The heterolayers show conductivities that are higher than those of pure CaF_2 or BaF_2 films (and also higher than that of $(\text{Ba,Ca})\text{F}_2$). The conductivity increases progressively with increasing interfacial density (decreasing period) in the period range 500 to 16 nm (here period refers to the thickness of CaF_2 plus BaF_2 unit layer). The change of conductivity with temperature is the same as expected for composite electrolytes³: at low temperature there is a conductivity enhancement characterized by an activation energy close to the migration enthalpy of the relevant carriers. In the heterogeneously doped solid electrolytes, this enhancement was attributed to a space-charge layer with an enhanced carrier concentration and an almost temperature-independent interfacial concentration.

Provided that the period is greater than ~ 100 nm, the conductivity increases linearly with N/L (Fig. 2); that is, approximately linearly with the number of heterojunctions N . The space-charge calculation^{3,11} (the details of which are beyond the scope of this Letter) uses published data on defect formation and mobilities^{18,19}, and sets the interfacial concentration equal to the inverse molar volume³; even though performed without any adjustable parameter, it yields surprisingly good agreement with the experiments (see dotted line in Figs 2 and 3). A perfect agreement is obtained when the data are background-corrected (that is, when contributions that are not due to the variable number of heterojunctions—such as contributions from bulk, surface and substrate interface—are accounted for), as shown by the bold lines in Figs 2 and 3. The evaluation of the data indicates that, at lower temperature, the space-charge layers in BaF_2 are determining the overall conductivity (compare the migration enthalpy of 0.75 eV). This is in agreement with the higher mobilities in BaF_2 compared with CaF_2 at these comparatively low temperatures. Owing to the similarity of the

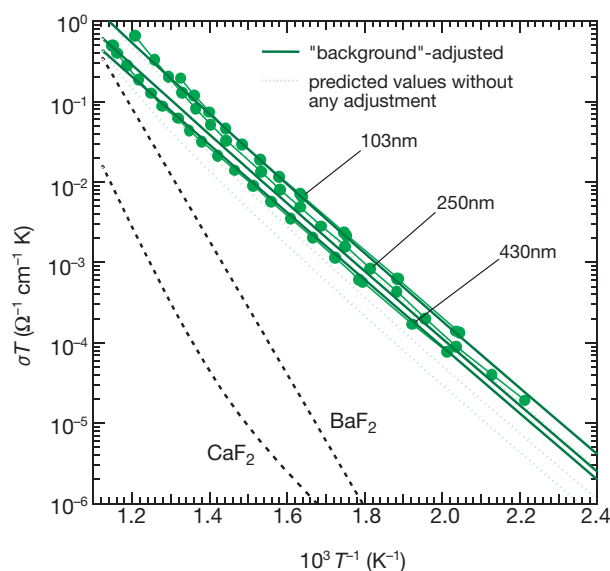


Figure 3 Comparison with space-charge calculations of the (parallel) conductivity results for the long-period films (green). Dotted lines, space-charge prediction with no background correction; solid lines, which perfectly agree with the experimental results (symbols), background-corrected. Dashed lines, the calculated intrinsic fluoride ion conductivities.

mobilities of interstitials and vacancies¹⁸ in BaF₂, further conclusions about the mechanism cannot be drawn at this stage (but see below).

Figure 2 also shows that the conductivities obtained for periods smaller than 100 nm (and greater than 15 nm) are distinctly larger than those estimated for semi-infinite boundary conditions. The critical spacing (half period) separating the regime of semi-infinite space-charge effects from the mesoscopic regime of overlapping space charges has been derived to be 4λ (ref. 12), λ being the Debye length. Indeed a Debye length of about 10 nm is, in view of the extrinsic–intrinsic transition of the pure materials, approximately what we had expected. Thus, the anomalous increase at periods less than 100 nm (that is, spacing less than 50 nm) strongly suggests the appearance of the searched-for nanosize effect. The nanosize factor derived in refs 3 and 12—which takes account of this additional increase by the overlap of space-charge layers—is comparable with the value obtained from our data. As long as the defect concentration in the film centre is still markedly smaller than at the interface, the nanosize factor is roughly given by $4\lambda/l$, where l is the interfacial spacing. For $l = 8$ nm and $\lambda \approx 10$ nm, we expect it to be about 5, which is to be compared with Fig. 2. Even the hyperbolic dependence on l is reflected by the data in Fig. 2 inset, yielding an improved and consistent Debye length of 15 nm. The detailed space-charge calculation is extremely involved and cannot be given here (N.S. and J.M., manuscript in preparation).

A further reduction of the period below about 15 nm yielded a conductivity decrease (see Fig. 2 inset). On the one hand, this is indeed theoretically expected for extremely thin films¹⁵ from the increased core–core interaction; on the other hand, however, the films are probably no longer continuous at these tiny thicknesses. Owing to the fact that the data start scattering (with respect to the l -dependence) at these extreme values, we now favour the second explanation.

We now consider the origin of the change in the activation energy of the short-period heterolayer samples at high temperatures in Fig. 1 (which differs in slope and onset temperature distinctly from the intrinsic bulk regime in the long-period heterolayer samples). An impurity effect would not be compatible with the space-charge considerations given here. Rather than being due to increasing conductivity of the core charge or to premelting effects, it seems that the more steeply activated CaF₂ space-charge conductivity determines the behaviour in this regime. Not only is the activation energy of 0.95 eV almost identical to the migration energy of F[−] interstitials in CaF₂ (ref. 19), but also the dominating contribution of the CaF₂ layers is an automatic result of the quantitative treatment (N.S. and J.M., manuscript in preparation). (We note that the nanosize factor approximated by $4\lambda/l$ does not depend on temperature as long as the bulk is extrinsic.) If this is the correct explanation, a transfer of

F[−] from BaF₂ to CaF₂ is indicated, rather than a bilateral segregation of F[−] to the interfacial core. The latter would also be not unexpected, as it would reduce the misfit strain. Whether this is indeed negligible—and to what degree the effects are influenced by interfacial mixing and to what extent further phenomena are involved—will need to be clarified in future studies.

Figure 4 displays the space-charge effects, which quantitatively account for the conductivity effects: it shows the semi-infinite space-charge situation as well as the mesoscopic situation in which there is no longer electro-neutral bulk. In the latter case, the layers have lost their individuality and an ensemble has formed with qualitatively new conductivity properties.

Our results open the way to the preparation of unusual artificial ion conductors and to the investigation of ionically conducting systems. Not only can interfacial effects be studied and used in high concentrations, but also mesoscopic phenomena can be explored—as shown here. The latter are important for potential applications in which a high conductivity normal to the interfaces is needed. Further systems of interest are: ...MX/MX... layers with intentionally modified boundaries; ...MX/A/MX... systems in which A represents a surface-active insulator; and heterolayers with mixed conductors in which both ionic and electronic conductivities are present. We expect the last of these systems to be of particular significance, because length scales determining size effects are different for electronic and ionic carriers. Whereas the effects discussed above are purely electrostatic in nature, we note that the standard term in the chemical potential of ionic defects changes if the structure changes¹⁵. For electrons, owing to their wave character and the related effective extension, such a size effect should occur on a much larger length scale. Like nano-electronics, the field of nano-ionics^{15,21–23} relies on the metastability of the phase distribution and is characterized by a high information content on a mesoscopic scale. Both fields deserve further intensive exploration by advanced methods such as the one presented here. □

Received 7 August; accepted 18 October 2000.

- Wagner, C. & Schottky, W. Theorie der geordneten Mischphasen. *Z. Phys. Chem. B* **11**, 163–210 (1930).
- Kudo, T. & Fueki, K. *Solid State Ionics* 1st edn (VCH, Weinheim, 1990).
- Maier, J. Ionic conduction in space charge regions. *Prog. Solid State Chem.* **23**, 171–263 (1995).
- Maier, J. Space charge regions in solid two phase systems and their conduction contribution - II: contact equilibrium at the interphase of two ionic conductors and the related conductivity effect. *Ber. Bunsenges. Phys. Chem.* **89**, 355–362 (1985).
- Maier, J. Defect chemistry and conductivity effects in heterogeneous solid electrolytes. *J. Electrochem. Soc.* **134**, 1524–1535 (1987).
- Liang, C. C. Conduction characteristics of the lithium iodide-aluminum oxide solid electrolytes. *J. Electrochem. Soc.* **120**, 1289–1292 (1973).
- Maier, J. Space charge regions in solid two phase systems and their conduction contribution - I. conductance enhancement in the system ionic conductor-‘inert’ phase and application on AgCl:Al₂O₃ and AgCl:SiO₂. *J. Phys. Chem. Solids* **46**, 309–320 (1985).
- Maier, J., Lauer, U. & Göpel, W. Gas-sensitivity of AgCl interfaces. *Solid State Ionics* **40/41**, 463–467 (1990).
- Saito, Y. & Maier, J. Ionic conductivity enhancement of the fluoride conductor CaF₂ by grain boundary activation using Lewis acids. *J. Electrochem. Soc.* **142**, 3078–3083 (1995).
- Shahi, K. & Wagner, J. B. Fast ion-transport in silver-halide solid-solutions and multiphase systems. *Appl. Phys. Lett.* **37**, 757–759 (1980).
- Maier, J. Enhancement of the ionic conductivity in solid-solid-dispersions by surface induced defects. *Ber. Bunsenges. Phys. Chem.* **88**, 1057–1062 (1984).
- Maier, J. Space charge regions in solid two phase systems and their conduction contribution - III: defect chemistry and ionic conductivity in thin films. *Solid State Ionics* **23**, 59–67 (1987).
- Maier, J. Point-defect thermodynamics and size effects. *Solid State Ionics* **131**, 13–22 (2000).
- Lee, J.-S., Adams, S. & Maier, J. Transport and phase transition characteristics in AgI : Al₂O₃ composite electrolytes - Evidence for a highly conducting 7-layer AgI polytype. *J. Electrochem. Soc.* **147**, 2407–2418 (2000).
- Maier, J. Point defect thermodynamics: Macro- vs. nanocrystals. *Electrochemistry* **68**, 395–402 (2000).
- Eberl, K., Petroff, P. M. & Demeester, P. (eds) in *Proc. NATO Advanced Res. Workshop* (Kluwer, Dordrecht, 1995).
- Farrow, R. F. C., Sullivan, P. W., Williams, G. M., Jones, G. R. & Cameron, D. C. MBE-grown fluoride films—a new class of epitaxial dielectrics. *J. Vac. Sci. Technol.* **19**, 415–420 (1981).
- Bollmann, W. Ionic conductivity of pure and doped BaF₂ crystals. *Phys. Status Solidi A* **18**, 313–321 (1973).
- Bollmann, W. & Reimann, R. Concentration and mobility of interstitial fluorine ions in CaF₂. *Phys. Status Solidi A* **16**, 187–196 (1973).
- Schoonman, J. in *Fast Ion Transport in Solids* (eds Vashishta, P., Mundy, V. & Shenoy, V) 631–636 (Elsevier, North-Holland, New York, 1979).



Figure 4 Comparison of conductivity profiles in the semi-infinite space-charge and mesoscale situations. The concentration or (parallel) conductivity profiles are sketched for the semi-infinite space-charge situation (period $> 8\lambda$, left), and for the mesoscale situation (period $< 8\lambda$, right) in which the space-charge regions overlap and bulk values are exceeded even in the centres of the individual layers.

21. Puin, W., Rodewald, S., Ramlau, R., Heitjans, P. & Maier, J. Local and overall ionic conductivity in nanocrystalline CaF_2 . *Solid State Ionics* **131**, 159–164 (2000).
22. Schoonman, J. Nanostructured materials in solid state ionics. *Solid State Ionics* **135–137**, 5–20 (2000).
23. Tuller, H. L. Ionic conduction in nanocrystalline materials. *Solid State Ionics* **131**, 143–157 (2000).

Acknowledgements

We thank W. Kussmaul for experimental assistance, G. Bilger and H. Kerber for the performance of SIMS and X-ray pole-figure measurements, and W. Dietsche for advice.

Correspondence and requests for materials should be addressed to J.M. (e-mail: weiglein@chemix.mpi-stuttgart.mpg.de) or to U.S. (e-mail: sata@chemix.mpi-stuttgart.mpg.de).

Pairing of isolated nucleic-acid bases in the absence of the DNA backbone

Eyal Nir*, Karl Kleinermanns† & Mattanjah S. de Vries*‡

* Department of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

† Institut für Physikalische Chemie und Elektrochemie, Heinrich Heine Universität, D-40225 Düsseldorf, Germany

‡ Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, USA

The two intertwined strands of DNA are held together through base pairing—the formation of hydrogen bonds between bases located opposite each other on the two strands. DNA replication and transcription involve the breaking and re-forming of these hydrogen bonds, but it is difficult to probe these processes directly. For example, conventional DNA spectroscopy^{1–3} is dominated by solvent interactions, crystal modes and collective modes of the DNA backbone; gas-phase studies, in contrast, can in principle measure interactions between individual molecules in

the absence of external effects, but require the vaporization of the interacting species without thermal degradation^{4–9}. Here we report the generation of gas-phase complexes comprising paired bases, and the spectroscopic characterization of the hydrogen bonding in isolated guanine–cytosine (G–C) and guanine–guanine (G–G) base pairs. We find that the gas-phase G–C base pair adopts a single configuration, which may be Watson–Crick, whereas G–G exists in two different configurations, and we see evidence for proton transfer in the G–C pair, an important step in radiation-induced DNA damage pathways¹⁰. Interactions between different bases and between bases and water molecules can also be characterized by our approach, providing stringent tests for high-level *ab initio* computations that aim to elucidate the fundamental aspects of nucleotide interactions^{11–13}.

In order to form isolated base pairs in the gas phase we desorb a mixture of neat bases from a graphite surface using 10-ns laser pulses at 1,064 nm and at 1 mJ cm^{−2}, followed by entrainment of the neutral, non-fragmented base molecules in a supersonic expansion of argon gas¹⁴. The expansion lowers the internal temperature of the molecules to a few tens of degrees Kelvin and induces cluster formation. In this way we have formed over 20 different base-pair combinations. By admixing water with the drive gas we can also form clusters of the bases with water molecules, simulating the solvent environment. We have formed such clusters ranging in size from one to fifty water molecules, attached to a single base molecule. We analyse the cold base pairs or clusters downstream with resonance-enhanced multiphoton ionization (REMPI): a photon from a tunable dye laser excites the molecules to the first excited singlet state, whereas a second photon ionizes only those molecules that are excited by the first. We detect the ions in a time-of-flight mass spectrometer. By varying the wavelength we obtain a mass-selected vibronic excitation spectrum. We obtained spectra of all individual bases, of guanine clusters with a single water molecule, and of a variety of base pairs (see Supplementary Information). In order to distinguish different cluster structures we apply spectral hole burning (SHB): while one laser is tuned over all transitions, a second laser, with a 30-ns delay and tuned permanently to a single

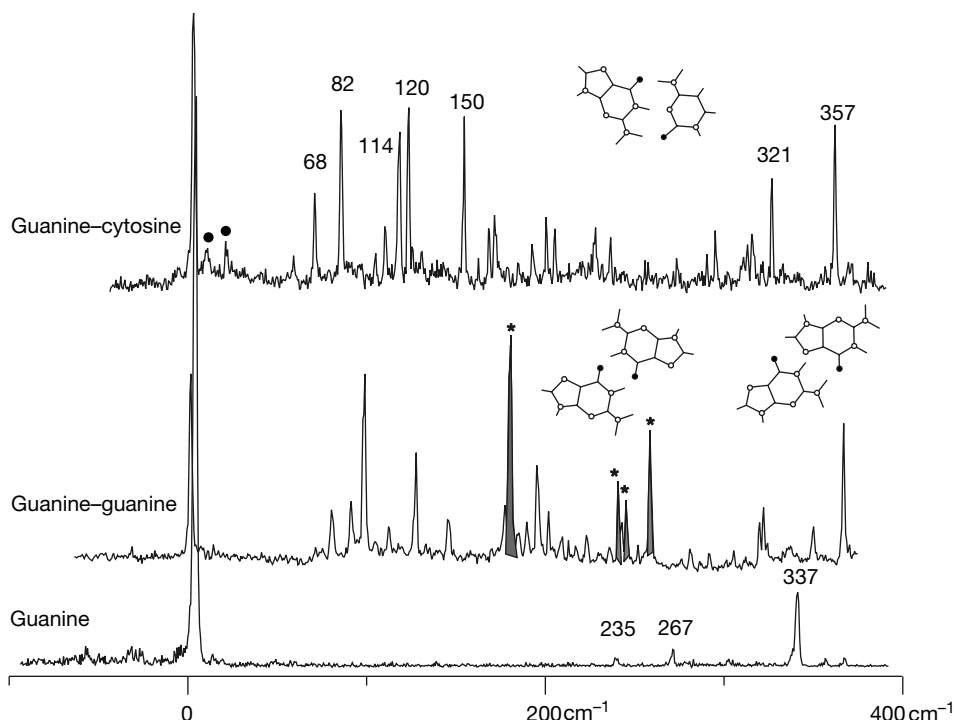


Figure 1 Resonance-enhanced multiphoton ionization (REMPI) spectrum of guanine and its dimers with guanine and cytosine. The spectra are offset such that the origins (0–0 transitions) are shown at the same position for all three species. See text for details.

transition, monitors ground-state depletion for a selected cluster species. Similarly, by using an infrared laser to deplete the ground state resonantly, we can obtain infrared spectra of selected species.

Figure 1 shows simultaneously recorded REMPI spectra for the G–C dimer, the G–G dimer and the G monomer, which are offset such that the origins (0–0 transitions) are shown at the same position. The G–C and G–G origins are blue-shifted by 446 cm^{-1} and by 235 cm^{-1} , respectively, with respect to the G origin at $32,868\text{ cm}^{-1}$. Neither the G nor the C chromophore of the dimers absorbs in the first 200 wavenumbers above the origin of the complex. The guanine spectrum exhibits its first significant vibronic activity at 235 cm^{-1} , whereas cytosine absorption is strongly blue-shifted to about $36,000\text{ cm}^{-1}$, as has been seen for uracil and thymine⁴. Therefore, when we excite the guanine chromophore in the G–G and G–C complexes, we can observe hydrogen bond vibrations built on the 0–0 transition without any interference. Indeed, all spectral lines in the first two hundred wavenumbers of these spectra can be assigned to hydrogen-bond vibrations.

The shaded peaks in the G–G spectrum are due to a different species, as determined by SHB, implying the presence of two different hydrogen-bonded G–G structures. In the case of G–C, SHB shows that all peaks belong to the same species; thus we observe one G–C structure only.

Table 1 summarizes the major lines for G–C. The frequencies of the less intense bands in the range $160\text{--}230\text{ cm}^{-1}$ all fit those of overtones and combination bands of the major lines. The bands at 321 and 357 cm^{-1} are the least energetic intramolecular G vibrations of G–C that we observed and probably correspond to in-plane ring deformations of G with symmetric and antisymmetric C–NH₂ and C=O bend motions, respectively¹³. The two low-frequency peaks in the spectrum, marked with circles, are absent when we use krypton as the drive gas instead of argon. This suggests that they may be due to hot bands or result from dissociating clusters with argon, rather than G–C vibrations.

The cluster comprised of one guanine base and one cytosine base has six intermolecular modes, reflecting the loss of three rotations and three vibrations upon complexation. We can describe the six possible modes as follows: (1) The butterfly motion, γ_s , with G and C as wings; (2) the out-of-plane torsion, τ ; and (3) the in-plane bending (gearing), δ , of G and C. These three motions can be correlated with monomer rotations approximately around the free short axis, the long axis and the axis perpendicular to the plane respectively. (4) The symmetric stretch, σ_s , and (5) the antisymmetric stretch, σ_{as} . These motions correlate with translations of the monomers parallel and perpendicular to the H-bond. (6) The displacements of G and C in opposite directions perpendicular to the molecular plane (alternating stairs), γ_{as} , correlating with monomer translations perpendicular to their planes.

In recent years, high-level *ab initio* calculations on DNA base

pairs have been carried out. The results of these calculations aid us with the interpretation of our spectra, but the experimental data also provide a firm benchmark for theory. In Table 1 we compare our experimental frequencies with the *ab initio* values calculated by refs 11–13 for the Watson–Crick (WC) configuration of this base pair. Agreement is particularly good for calculations at the HF/6-31G(d) and the HF/6-31G(d,p) levels, which are known to yield relatively reliable harmonic intermolecular frequencies without the need to resort to scaling factors^{15,16}. Deviations due to anharmonicity should be important only for out-of-plane vibrations, the first two of which are missing in our G–C spectrum owing to their low transition strength. It seems reasonable to compare ground-state calculations and excited-state measurements, because the guanine origin intensity is an order of magnitude larger than that of its vibronic transitions, which suggests that there is little difference between the geometry of guanine in the ground, and the excited state. In order to assess the extent to which the agreement with calculations for the WC structure constitutes a unique fit, we have also calculated the vibrational frequencies for eight other structures, obtained by minimizing different starting configurations. These computations, at the HF 6-31G(d,p) level, give results for the WC structure identical to those of ref. 12 and agree rather less well with the experimental data for all other structures. Further determination for the gas-phase G–C structure will follow from infrared spectral hole burning on the G origin of the pair, for which we can observe N–H and O–H stretch frequencies, as well as their shifts upon complexation (see the Supplementary Information for IR hole burning results on the isolated G). We have also studied clusters between guanine and methyl-substituted cytosines. Clusters with 5-methylC show spectra identical to those of the unsubstituted ones, while clusters with 3-methylC show a very different spectrum. These results are also consistent with a WC structure for the unsubstituted G–C pairs. Further support for this conclusions might be derived from band-contour and isotope-shift measurements.

The REMPI spectrum of the G–G base pair likewise shows low-frequency vibrations in the range up to 200 cm^{-1} , which can again be ascribed to intermolecular vibrations and their overtones and combination bands. Spectral hole burning reveals the existence of two sets of peaks in this spectrum (distinguished by peak shading and asterisks in Fig. 1), implying two different cluster structures. *Ab initio* calculations of the G–G structure at different levels of theory agree that the structures displayed in Fig. 1 are the two most stable ones^{17,18}. When both monomer parts in the dimer are identical, only one electronic spectrum is expected, whereas exciton splitting may lead to a second electronic band system. A further analysis, which awaits more detailed calculations for this base pair, should thus allow us to distinguish between the symmetric and the non-symmetric structure.

Table 1 Intermolecular vibrational frequencies of the G–C base pair in the S₁ state

Frequency* (cm ⁻¹)	Relative intensity†	Label	Description	<i>Ab initio</i>		
				HF/6-31G(d)‡	HF/6-31G(d,p)§	BP86/6-3111G
		γ_s	Symmetric out-of-plane bending (butterfly)	20	25	29
		τ	Torsion	32	34	44
67.5	0.30	γ_{as}	Antisymmetric out-of-plane bending	66	69	85
81.8	0.64	δ ¶	In-plane bending	84	84	105
106.4	0.17	$\gamma_s + \delta/\gamma_{as} + \tau$				
114.4	0.54	σ_{as} ¶	Antisymmetric stretching	117	116	145
119.5	0.65	σ_s	Symmetric stretching	125	123	150
150.4	0.62	$\gamma_{as} + \sigma_{as}/\sigma_s$ #				

* Relative to the electronic origin at $33,313\text{ cm}^{-1}$; experimental uncertainty is $\pm 0.5\text{ cm}^{-1}$.

† Relative to origin intensity.

‡ Ref. 3.

§ Ref. 4.

|| Ref. 5.

¶ The two normal mode motions are very similar and can also be regarded as rotations of G (or C) about the axis perpendicular to their planes with the other cluster moiety, C (or G) performing an opposite, smaller-amplitude rotation to conserve the centre of mass.

The solidus indicates alternative assignments.

Experiments using laser desorption from an equimolar mixture of guanine, adenine, cytosine, thymine and uracil yielded mass peaks for all combinations of G with the other bases and clusters comprised of two, three and four bases (see Supplementary Information). The abundance ratios depend on wavelength because ionization profiles and ionization efficiencies differ. *Ab initio* calculations at different levels of theory agree that the G–C Watson–Crick base pair with three nearly linear hydrogen bonds and the symmetrically hydrogen-bonded G–G base pair are the most stable base pairs¹⁹. These two base pairs indeed appear to form preferentially in our system, but we cannot derive absolute association constants because jet expansion clusters are not formed under thermal equilibrium conditions.

Proton transfer between bases in a base-pair radical cation and subsequent reactions in the nucleotides are assumed to be a major source of damage of DNA by ionizing radiation¹⁰. We observed the electronic origin band of G–C also as a spectral peak at the mass of $(C + H)^+$ ($m/e = 112$). This implies a fast $G \rightarrow C$ proton transfer after excitation, followed by dissociation into the $(C + H)^+$ ion and either the $(G - H)$ radical or the $(G - H)^-$ negative ion. The hydrogen transfer can take place either in the S_1 excited state or in the ionic ground state. The measurements were performed at rather low laser fluence ($\sim 100 \mu J$, 10-ns pulse width, 2-mm beam diameter) such that absorption of a further photon by the $(G + C)^+$ ion is unlikely. The total excitation energy available from two photons at the resonant wavelength is 8.26 eV ($2 \times 33313 \text{ cm}^{-1}$). This is very close to the vertical ionization potential of guanine²⁰ and is probably insufficient to lead to $G^+ + C$ as products. The fact that we do observe dissociation of the ion with $(C + H)^+$ as a product channel at this energy, but that dissociation into $G^+ + C$ does not occur, suggests that proton transfer weakens the bonding in the complex ion.

We have generated gas-phase complexes comprised of selected DNA bases whose interaction can be probed free of interactions with other bases or solvent molecules. This approach allows the generation of gas-phase complexes containing a single base and water molecules as well, and the use of mass-selected detection should therefore make it possible to elucidate the influence of the solvent in detail, one water molecule at a time²¹. Finally, as the electronic ground state of water clusters is experimentally accessible by using infrared-ultraviolet double resonance spectroscopy, we expect that detailed comparisons with theoretical predictions will be possible. □

Received 27 August 1999; accepted 9 November 2000.

1. Fodor, S. P. *et al.* H/Sub 2/Raman-shifted YAG laser ultraviolet raman spectrometer operating at wavelengths down to 184 nm. *J. Raman Spectrosc.* **17**, 471–475 (1986).
2. Urabe, H. *et al.* Collective vibrational modes in molecular assembly of DNA and its application to biological systems. Low frequency raman spectroscopy. *J. Chem. Phys.* **82**, 531–535 (1985).
3. Cocco, S. & Monasson, R. Theoretical study of collective modes in DNA at ambient temperature. *J. Chem. Phys.* **112**, 10017–10033 (2000).
4. Brady, B. B., Peteau, L. A. & Levy, D. H. The electronic spectra of the pyrimidine bases uracil and thymine in a supersonic molecular beam. *Chem. Phys. Lett.* **147**, 538–543 (1988).
5. Viant, M. R. *et al.* Infrared laser spectroscopy of uracil in a pulsed slit jet. *J. Chem. Phys.* **103**, 9502–9505 (1995).
6. Nir, E. *et al.* REMPI spectroscopy of jet cooled guanine. *J. Am. Chem. Soc.* **121**, 4896–4897 (1999).
7. Yanson, I. K., Teplitsky, A. B. & Sukhodub, L. F. Molecular interactions in nucleic acids. *Biopolymers* **18**, 1149–1170 (1997).
8. Dey, M. *et al.* Base pair formation of free nucleobases and mononucleosides in the gas phase. *J. Am. Chem. Soc.* **116**, 9211–9215 (1994).
9. Dey, M., Grottemeyer, J. & Schlag, E. W. Pair formation of free nucleobases and mononucleosides in the gas-phase. *Z. Naturforsch.* **49A**, 776–784 (1994).
10. Hutter, M. & Clark, T. On the enhanced stability of the guanine-cytosine base-pair radical cation. *J. Am. Chem. Soc.* **118**, 7574–7577 (1996).
11. Florian, J., Leszczynski, J., Johnson, B. G. On the intermolecular vibrational modes of the guanine-cytosine, adenine-thymine and formamide-formamide H-bonded dimers. *J. Mol. Struct.* **349**, 421–426 (1995).
12. Shishkin, O., Sponer, J., Hobza, P. Intermolecular flexibility of DNA bases in adenine-thymine and guanine-cytosine Watson-Crick base pairs. *J. Mol. Struct.* **477**, 15–21 (1999).
13. Santamaria, R. *et al.* Vibrational spectra of nucleic acid bases and their Watson-Crick pair complexes. *J. Comput. Chem.* **20**, 511–530 (1999).
14. Meijer, G. *et al.* Laser desorption jet-cooling of organic molecules. Cooling characteristics and detection sensitivity. *Appl. Phys. B* **51**, 395–403 (1990).

15. Schutz, M., Burgi, T. & Leutwyler, S. Intermolecular bonding and vibrations of phenol.H₂O (D₂O). *J. Chem. Phys.* **98**, 3763–3777 (1993).
16. Schmitt, M., Jacoby, C. & Kleinermanns, K. Torsional splitting of the intermolecular vibrations of phenol (H₂O)1 and its deuterated isotopomers. *J. Chem. Phys.* **108**, 4486–4495 (1997).
17. Sponer, J., Leszczynski, J. & Hobza, P. Structures and energies of hydrogen-bonded DNA base pairs. A nonempirical study with inclusion of electron correlation. *J. Phys. Chem.* **100**, 1965–1974 (1995).
18. Hobza, P. *et al.* Performance of empirical potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), semiempirical quantum chemical methods (AM1, MNDO/M, PM3), and *ab initio* Hartree-Fock method for interaction of DNA bases: Comparison with nonempirical beyond Hartree-Fock results. *J. Comput. Chem.* **18**, 1136–1150 (1997).
19. Hobza, P. & Sponer, J. Thermodynamic characteristics for the formation of H-bonded DNA base pairs. *Chem. Phys. Lett.* **261**, 379–384 (1996).
20. Colson, A. *et al.* *Ab initio* molecular orbital calculations on DNA base pair radical ions: Effect of base pairing on proton-transfer energies, electron affinities, and ionization potentials. *J. Phys. Chem.* **96**, 9787–9794 (1992).
21. Janzen, C. *et al.* Structure and vibrations of Phenol(H₂O)_{7,8} studied by infrared-ultraviolet and ultraviolet-ultraviolet double-resonance spectroscopy and *ab initio* theory. *J. Chem. Phys.* **110**, 9898–9907 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work has been supported by the Deutsche Forschungsgemeinschaft and United States-Israel Binational Science Foundation.

Correspondence and requests for materials should be addressed to M.S.d.V. (e-mail: devries@chem.ucsb.edu).

The influence of rivers on marine boron isotopes and implications for reconstructing past ocean pH

D. Lemarchand, J. Gaillardet, É. Lewin & C. J. Allègre

Laboratoire de Géochimie et Cosmochimie, Institut de Physique du globe de Paris, 4 place Jussieu, 75252 Paris Cedex 05

Ocean pH is particularly sensitive to atmospheric carbon dioxide content^{1–3}. Records of ocean pH can therefore be used to estimate past atmospheric carbon dioxide concentrations. The isotopic composition of boron ($\delta^{11}\text{B}$) contained in the carbonate shells of marine organisms varies according to pH, from which ocean pH can be reconstructed^{4–11}. This requires independent estimates of the $\delta^{11}\text{B}$ of dissolved boron in sea water through time. The marine $\delta^{11}\text{B}$ budget, however, is still largely unconstrained. Here we show that, by incorporating the global flux of riverine boron (as estimated from $\delta^{11}\text{B}$ measurements in 22 of the world's main rivers), the marine boron isotope budget can be balanced. We also derive ocean $\delta^{11}\text{B}$ budgets for the past 120 Myr. Estimated isotope compositions of boron in sea water show a remarkable consistency with records of $\delta^{11}\text{B}$ in foraminiferal carbonates^{9–11}, suggesting that foraminifera $\delta^{11}\text{B}$ records may in part reflect changes in the marine boron isotope budget rather than changes in ocean pH over the Cenozoic era.

The present-day concentration of boron in the oceans and its isotopic composition ($\delta^{11}\text{B}_{\text{sw}}$) are both uniform, with values of 4.5 p.p.m. and +39.6‰, respectively ($\delta^{11}\text{B}$ is defined in Table 1). Although a number of previous investigations have led to the identification of the main processes affecting boron in the ocean, the oceanic cycle of boron and its secular evolution are still not known. The sinks of boron in the ocean are divided into three classes: uptake during low-temperature weathering of the oceanic crust^{12,13}, adsorption on clastic sediments^{14,15} and coprecipitation in carbonates⁴. To balance these outputs from the oceans, three sources have been proposed: rivers, hydrothermal vents and fluid expelled

from accretionary prisms in subduction zones. However, to date, these fluxes have not resulted in a balanced oceanic mass budget¹³.

Although assessment of the input of boron from the continents and its isotopic composition in river waters is of crucial interest, studies of boron systematics in large rivers have been hampered by analytical difficulties and therefore still remain an outstanding problem. The few data available for the isotopic composition of boron in large rivers range from +7‰ and +21‰ (refs 16, 17). Following a recently developed chemical procedure¹⁸, we have focused on the analysis of boron in 22 of the world's largest rivers. This offers the possibility of getting a reasonable estimate of the boron riverine discharge to the ocean by integrating over large continental surfaces, and should therefore maximize the likelihood of sampling. The analysed samples represent 39% of the total world discharge of water and cover a large range of climate conditions, from arid to tropical and from equatorial to subarctic. Results are reported in Table 1, where rivers are ranked by boron flux.

The Amazon contributes a significant flux of boron (about 10% of the total boron input to the ocean) owing to its extremely large water discharge. Large variations among the rivers are observed both for dissolved boron concentrations, [B], and isotopic compositions. Concentrations range between 1 p.p.b. for the Fraser River and 201 p.p.b. for the Huanghe River, with 85% of the analysed rivers having boron contents less than 25 p.p.b. Boron isotope compositions range between -6‰ for the Brahmaputra River and +42.8‰ for the Maroni River, with 75% of the riverine $\delta^{11}\text{B}$ between 2‰ and 20‰. We investigated seasonal variations for several rivers by sampling during high and low water stages. The Amazon and Changjiang rivers show constant [B] and $\delta^{11}\text{B}$ throughout hydrological cycles, whereas the Huanghe and Xijiang rivers show seasonal variations on the order of 30% and 10‰ for [B] and $\delta^{11}\text{B}$, respectively (see Table 1). As corrections for potential seasonal changes are impossible to apply to other samples, we used as many high-water-stage samples as possible for the total flux calculation in order to minimize uncertainties. Assuming that the analysed samples are representative of the world average, we calculate a discharge-weighted mean river concentration of 10.2 p.p.b. and a $\delta^{11}\text{B}$ of 10‰. We then estimate the total continental boron input to

the ocean as $38 \times 10^{10} \text{ g B yr}^{-1}$, by extrapolating from our data to the world-total water discharge. (Errors are difficult to estimate, but we assume them to be on the order of 20%.) Such a flux is consistent with the pioneering estimate of ref. 19.

In fact, the flux of B that we estimate here includes seawater-derived atmospheric boron (so-called cyclic boron) and the anthropogenic contribution; neither of these contributions should be taken into account for the long-term oceanic budget. We note that the contribution of atmospheric B to rivers, based on the determination of the B content of rainwaters, has been shown to be negligible for Himalayan rivers¹⁷. Continuing research on the world-wide systematics of rainwater will provide a better evaluation of cyclic boron. As boron is used in detergents and soaps, riverine boron may be affected by anthropogenic materials²⁰. From the yearly production of boron, and the proportion of boron used in detergents and soaps²¹ and therefore potentially released into rivers, we estimate the anthropogenic discharge flux at $2.5 \times 10^{10} \text{ g B yr}^{-1}$, which represents less than 7% of the global natural flux of boron estimated here (within the error estimated above).

The riverine flux of B that we report here, together with the other inputs and outputs taken from literature, leads to a balanced oceanic mass budget for boron (Table 2). The present study

Table 2 Boron oceanic budget

	Boron flux ($10^{10} \text{ g yr}^{-1}$)	$\delta^{11}\text{B}$ (‰)	$\Delta_{\text{out-sw}}$ (‰)	Ref.
Ocean outputs				
Oceanic crust alteration	27	4	-36	13
Adsorption on sediment	13	15 ± 1	-25	15*
Coprecipitation in carbonates	6	20 ± 5	-20	4
Total	46	9.2	-30	
Ocean inputs				
Weathering	38	10		This study
Hydrothermal	4	6.5 ± 8		13, 27
Fluid expelled from accretionary prisms	2	25 ± 5		13, 28
Total	44	10.4		

* With the estimation of sediment discharge to the ocean from ref. 29.

$\Delta_{\text{out-sw}} = \delta^{11}\text{B} - \delta^{11}\text{B}_{\text{sw}}$.

Table 1 Synthesis of the data on water samples

River and sampling location	Date of sampling	Water stage*	Discharge ($\text{km}^3 \text{ yr}^{-1}$)†	[B]‡ (p.p.b.)	Boron flux (tons yr^{-1})	$\delta^{11}\text{B}$ § (‰)	Error $\pm 2\sigma$
Amazon at mouth	Oct. 98	H	6590	6.1	40,199	15.20	0.27
Mississippi II		L	529	37.8	19,996	14.0	1
Brahmaputra at Aricha Ghat	Aug. 99	H	510	20.9	10,659	-5.94	0.15
Changjiang at Nanjing	Aug. 94	H	928	11	10,208	4.04	0.16
Changjiang at Nanjing	Nov. 93	L		13.7		4.16	0.09
Gange at Rajbari	Aug. 99	H	493	17.8	8,775	14.36	0.18
St-Lawrence at Cornwall	Sep. 95	M	337	24.8	8,358	8.22	0.13
HuangHe at Jinan	Aug. 94	H	41	201	8,241	3.10	0.11
HuangHe at Jinan	Nov. 93	L		131		10.96	0.19
Irrawaddy at Sitsayan	Feb. 99	L	486	16.9	8,213	13.49	0.13
Salween at Bawthabya	Feb. 99	L	211	37.5	7,913	-3.11	0.15
Mekong at Vientiane	Aug. 92	H	467	14.7	6,865	2.31	0.24
Congo at Brazzaville	Nov. 89	H	1200	3.1	3,720	28.37	0.21
Mackenzie at Tsiigehtchic	Aug. 96	H	308	11.8	3,634	14.94	0.17
Orinoco at Ciudad Bolivar	Jun. 98	H	1135	2.4	2,724	10.76	0.55
Lena at mouth	Aug. 98	H	525	4.7	2,468	8.57	0.15
Colorado at Needles	Jun. 97	L	18.5	127.9	2,366	11.56	0.18
Xijiang at Zhaoqing	Aug. 94	H	363	5.4	1,960	18.23	0.14
Xijiang at Zhaoqing	Nov. 93	L		7.4		6.53	0.19
Niger at Agenebode	Nov. 95	H	154	3.2	493	35.22	0.18
Narmada at Hoshangabad	Jun. 98	H	39	11.6	452	20.95	0.17
Tapti at Bhusawal	Jun. 98	H	18	24.3	437	32.56	0.36
Seine at Paris	May. 97	M	12.9	25	323	3.42	0.47
Maroni (French Guyana)	-	H	57.4	4.5	258	42.76	0.16
Fraser at Prince George	Jun. 99	H	112	1.0	112	11.85	0.33

* Water stage: H, high; L, low; M, median.

† Water discharges correspond to the mean annual discharge.

‡ The boron concentrations were determined by isotope dilution with precision better than 1%.

§ Boron isotope ratios were measured by positive TIMS using the $\text{Cs}_2\text{B}_2\text{O}_7$ method with graphite and mannitol; analytical reproducibility, 0.3‰ ($\pm 2\sigma$). $\delta^{11}\text{B} = \{[(^{11}\text{B}/^{10}\text{B})_{\text{sample}} / (^{11}\text{B}/^{10}\text{B})_{\text{standard}}] - 1\}$ where standard is NBS SRM 951.

|| Data from ref. 16.

emphasizes the importance of the continental flux in the oceanic B cycle. The fluxes listed in Table 2 allow the calculation of a 14-Myr residence time for boron in the oceans. This value is lower than the 20 Myr estimated in ref. 12. In addition, we have calculated the flux-weighted isotopic composition for sources and sinks of boron in the ocean. The averaged $\delta^{11}\text{B}$ of the inputs is 10.4‰ and the averaged output $\delta^{11}\text{B}$ is 9.2‰. A rough error calculation based only on the isotopic data dispersion for each flux leads to an uncertainty of at least 2‰ in both averaged fluxes. These results imply that the ocean is at, or close to, steady state with respect to boron isotopes. One related question then arises: could the boron isotope composition of sea water have changed significantly over geological timescales?

Potential causes for change in the boron isotope composition of the oceans with time are the variations of the relative proportions of each output and long-term changes of the weathering-derived input. We present a simple model for boron in the oceans, and

use it to estimate the possible changes in global boron content and the boron isotope composition of the oceans. The model can be summarized by two simple equations, as follows. Boron content is given by:

$$\frac{dB_{\text{sw}}}{dt} = \sum \Phi_{\text{in}} - \sum \Phi_{\text{out}}$$

and the boron isotope budget is given by:

$$\frac{dR_{\text{sw}}}{dt} = \sum \frac{\Phi_{\text{in}}}{B_{\text{sw}}} \times \frac{1 + R_{\text{sw}}}{1 + R_{\text{in}}} \times (R_{\text{in}} - R_{\text{sw}}) - \sum \frac{\Phi_{\text{out}}}{B_{\text{sw}}} \times \frac{1 + R_{\text{sw}}}{1 + R_{\text{out}}} \times (R_{\text{out}} - R_{\text{sw}})$$

Here B_{sw} and R_{sw} are respectively the total amount of boron and the $^{11}\text{B}/^{10}\text{B}$ ratio in seawater; Φ_{in} , Φ_{out} and R_{in} , R_{out} are respectively the fluxes and the $^{11}\text{B}/^{10}\text{B}$ ratio of the inputs and outputs. In this model, the B uptake of each sink is proportional to the total amount of boron in sea water, and the fractionation factors are taken to be constant for each process.

Several recent studies^{22–24}, covering the past 140 Myr, offer good estimates of the temporal variations of sea-floor generation, sediment production and carbonate precipitation (Fig. 1); we used these estimates to derive the associated boron fluxes. As the secular evolution of the riverine input is poorly constrained, we consider two scenarios: constant riverine flux, and changes of runoff with time.

In the first scenario, we assume a constant continental input of boron equal to the modern value. The results show a continuous increase of $\delta^{11}\text{B}_{\text{sw}}$ over the past 60 Myr (solid line in Fig. 2a). The behaviour of the boron isotope composition over the past 20 Myr

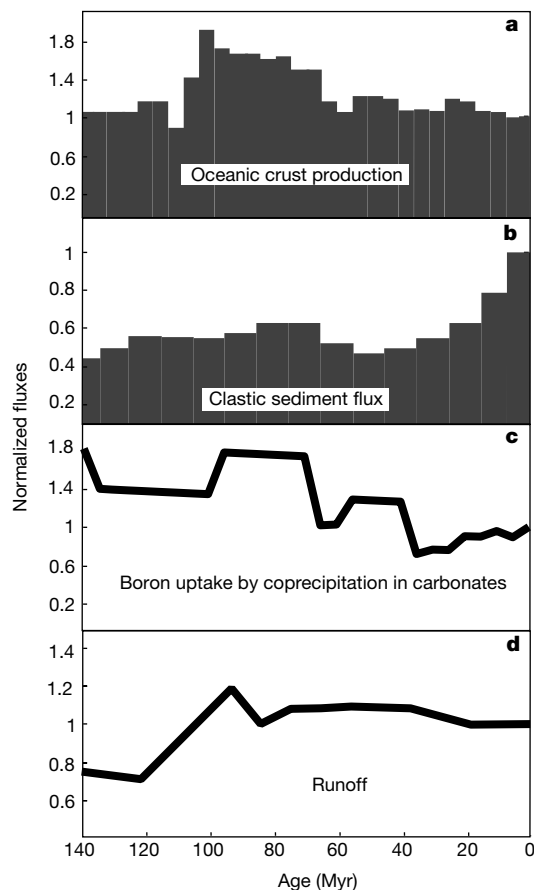


Figure 1 Estimation of the changes over the past 140 Myr of the main processes controlling the boron budget in the ocean. The variations are normalized to modern values. **a**, Oceanic crust production²². **b**, Clastic sediment flux²³. **c**, Boron uptake by coprecipitation in carbonate^{4,24}. **d**, Runoff variations²⁵. The magnitude of the boron sink during the low-temperature weathering of the oceanic crust is calculated from the total amount of fresh basalt exposed to sea water and assumed to be directly related to the spreading rate of the oceanic crust. Similarly, adsorption on sediments is calculated from the variations of the clastic sediment flux, and coprecipitation in carbonates is derived from the total carbonate deposit. The evolution of pelagic and platform carbonate deposits over the past 140 Myr has been reported²⁴. Because the magnitude of the boron sink is different between calcite and aragonite, we have calculated the associated boron flux taking into account the rate of boron sink for corals and forams as defined in ref. 4. Additionally, hydrothermal input and fluid expelled from accretionary prisms in subduction zones are assumed to be proportional to the oceanic crust spreading rate. We assume that both the partition coefficient and the fractionation factors are independent of the pH, temperature and salinity of the ocean.

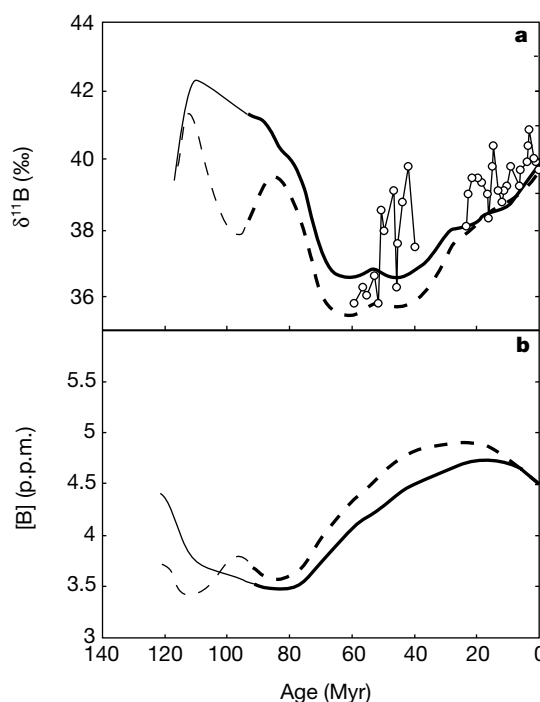


Figure 2 Models of the secular evolution of boron in the ocean. We consider two different histories through time for the continental input of boron. Solid line, constant riverine input; dashed line, riverine input following the runoff changes. The first 20 Myr (that is, from 140 to 120 Myr ago) of the models have not been represented because of their strong dependence on the initial values, which are arbitrary. **a**, Models of the isotopic composition of boron in the oceans through time. Open circles, interpreted data from ref. 11. **b**, Models of the boron concentration in the oceans.

seems to be dominated by the change of clastic sediment flux. The large enhancement of the spreading rate of the oceanic crust during the Cretaceous period leads to a significant increase of $\delta^{11}\text{B}_{\text{sw}}$ consistent with a decrease in boron concentration (Fig. 2b). The second scenario takes into account the variations of runoff through time proposed in ref. 25 (see Fig. 1d). This leads to a secular evolution of boron isotopes in the oceans (dashed line in Fig. 2a) similar to that observed in the first model. Although the variations of runoff were relatively large (on the order of 25%), such long-term variations of runoff do not appear to play a significant role in the boron oceanic budget. These two models lead to the calculation of a typical rate of change of boron isotope composition in the oceans of about 0.1‰ Myr^{-1} . We conclude that the determination of boron isotope ratio having analytical errors on the order of 0.3‰ , the hypothesis of $\delta^{11}\text{B}_{\text{sw}}$ remaining strictly constant with time is not tenable longer than 3 Myr before present.

We now compare these modelled evolutions with boron isotope data in carbonates. Palmer *et al.*⁹ analysed boron isotopes in species of planktonic foraminifera over the past 16 Myr. They discussed the measured decrease with age of the isotopic compositions of the foraminifera in the context of three hypothesis: variation of the oceanic pH value, variation of the temperature of the Pacific Ocean mixed layer, and variations in $\delta^{11}\text{B}_{\text{sw}}$. Applying the third hypothesis, they found an apparent decrease in $\delta^{11}\text{B}_{\text{sw}}$ of approximately 2‰, 16 Myr ago. In addition, Pearson and Palmer^{10,11} deduced palaeo-pH values of sea water from boron isotope compositions of planktonic foraminifera from the middle Eocene epoch and the Cenozoic era. Assuming that these data reflect variations in $\delta^{11}\text{B}_{\text{sw}}$ rather than pH, $\delta^{11}\text{B}_{\text{sw}}$ values 2–4‰ lower than modern values may be derived (Fig. 2a). Such a decrease is consistent with our scenarios, tending to indicate that the pH values of the oceans did not change significantly on this timescale. Following the GEOCARB model²⁶—where important changes of atmospheric $p\text{CO}_2$ are expected over the Cenozoic era—this result suggests that buffering mechanisms exist in the oceans that maintain the pH at a roughly constant value on geological timescales.

Accordingly, boron isotope composition remains a potential proxy for oceanic palaeo-pH reconstruction during the last glacial–interglacial period, but precise determination of the secular evolution of $\delta^{11}\text{B}_{\text{sw}}$ is required for greater ages. More data needs to be collected to enable the reconstruction of ancient oceanic $\delta^{11}\text{B}$ values over geological timescales. □

Received 20 June; accepted 6 November 2000.

- Berner, R. A. The rise of plants and their effect on weathering and atmospheric CO_2 . *Science* **276**, 544–546 (1997).
- Goddéris, Y. & François, L. M. Balancing the Cenozoic carbon and alkalinity cycles: Constraints from isotopic records. *Geophys. Res. Lett.* **23**, 3743–3746 (1996).
- McCauley, S. E. & DePaolo, D. J. *Tectonic Uplift and Climate Change* (Plenum, New York, 1997).
- Vengosh, A., Kolodny, Y., Starinsky, A., Chivas, A. R. & McCulloch, M. T. Coprecipitation and isotopic fractionation of boron in modern biogenic carbonates. *Geochim. Cosmochim. Acta* **55**, 2901–2910 (1991).
- Hemming, N. G. & Hanson, G. N. Boron isotopic composition and concentration in modern marine carbonates. *Geochim. Cosmochim. Acta* **56**, 537–543 (1992).
- Spivack, A. J., You, C. F. & Smith, J. Foraminiferal boron isotope ratios as a proxy for surface ocean pH over the past 21 Myr. *Nature* **363**, 149–151 (1993).
- Gaillardet, J. & Allégre, C. J. Boron isotopic compositions of corals: seawater or diagenesis record? *Earth Planet. Sci. Lett.* **136**, 665–676 (1995).
- Sanyal, A., Hemming, N. G., Hanson, G. N. & Broecker, W. S. Evidence for a higher pH in the glacial ocean from boron isotopes in foraminifera. *Nature* **373**, 234–236 (1995).
- Palmer, M. R., Pearson, P. N. & Cobb, S. J. Reconstructing past ocean pH–depth profiles. *Science* **282**, 1468–1471 (1998).
- Pearson, P. N. & Palmer, M. R. Middle Eocene seawater pH and atmospheric carbon dioxide concentrations. *Science* **284**, 1824–1826 (1999).
- Pearson, P. N. & Palmer, M. R. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature* **406**, 695–699 (2000).
- Spivack, A. J. & Edmond, J. M. Boron isotope exchange between seawater and the oceanic crust. *Geochim. Cosmochim. Acta* **51**, 1033–1043 (1987).
- Smith, H. J., Spivack, A. J., Staudigel, H. & Hart, S. R. The boron isotopic composition of altered oceanic crust. *Chem. Geol.* **126**, 119–135 (1995).
- Schwarz, H. P., Agyei, E. K. & McMullen, C. C. Boron isotopic fractionation during clay adsorption from sea-water. *Earth Planet. Sci. Lett.* **6**, 1–5 (1969).

- Spivack, A. J., Palmer, M. R. & Edmond, J. M. The sedimentary cycle of the boron isotopes. *Geochim. Cosmochim. Acta* **51**, 1939–1949 (1987).
- Spivack, A. J. *Boron Isotope Geochemistry*. Thesis, Massachusetts Institute of Technology/Woods Hole Oceanographic Institute (1986).
- Rose, E. F., Chaussidon, M. & France-Lanord, C. Fractionation of boron isotopes during erosion processes: the example of Himalayan rivers. *Geochim. Cosmochim. Acta* **64**, 397–408 (2000).
- Lemarchand, D., Gaillardet, J., Göpel, C. & Manhès, G. An optimized procedure for boron separation and mass spectrometry for river water samples. *Chem. Geol.* (submitted).
- Harris, R. C. Boron regulation in the oceans. *Nature* **223**, 290–291 (1969).
- Vengosh, A., Spivack, A. J., Artzi, Y. & Ayalon, A. Geochemical and boron, strontium, and oxygen isotopic constraints on the origin of the salinity in groundwater from the Mediterranean coast of Israel. *Wat. Resour. Res.* **35**, 1877–1894 (1998).
- Reimann, C. & Caritat, P. *Chemical Elements in the Environment* (Springer, Berlin, 1998).
- Larson, R. L. The Mid-Cretaceous superplume episode. *Sci. Am.* **272**, 66–70 (1995).
- Goddéris, Y. *Modélisation de l'Évolution Cénozoïque des Cycles Biogéochimiques: Impact de l'Orogénèse Himalayenne*. Thesis, Univ. Liège (1997).
- Opdyke, B. N. & Wilkinson, B. H. Surface area control of shallow cratonic to deep marine carbonate accumulation. *Paleoceanography* **3**, 685–703 (1988).
- Gibbs, M. T., Bluth, G. J. S., Fawcett, P. J. & Kump, L. R. Global chemical erosion over the last 250 Myr: variations due to changes in paleogeography, paleoclimate, and paleogeology. *Am. J. Sci.* **299**, 611–651 (1999).
- Berner, R. A. GEOCARB II: A revised model of atmospheric CO_2 over Phanerozoic time. *Am. J. Sci.* **294**, 56–91 (1994).
- You, C. F., Spivack, A. J., Smith, J. H. & Gieskes, J. M. Mobilization of boron in convergent margins: implications for the boron geochemical cycle. *Geology* **21**, 207–210 (1993).
- You, C. F., Spivack, A. J., Gieskes, J. M., Rosenbauer, R. & Bishoff, J. L. Experimental study of boron geochemistry: Implications for fluid processes in subduction zones. *Geochim. Cosmochim. Acta* **59**, 2435–2442 (1995).
- Milliman, J. D. & Syvitski, J. P. M. Geomorphic/tectonic control of sediment discharge to the ocean: the importance of small mountainous rivers. *J. Geol.* **100**, 525–544 (1992).

Acknowledgements

We thank R. A. Berner, G. N. Hemming, A. J. Spivack, E. Young and B. Bourdon for comments on the manuscript. We also thank the following for sampling facilities: T. Allard, M. Benedetti, M. Blasco, F. Bodet, G. J. Chakrapani, B. Dupré, C. Gariépi, W. Li, P. Negrel, V. Rachold, V. Subramanian and J. Zhao.

Correspondence and requests for materials should be addressed to D.L. (e-mail: lemarcha@ipgp.jussieu.fr).

Evidence from episodic seamount volcanism for pulsing of the Iceland plume in the past 70 Myr

J. M. O'Connor*, P. Stoffers†, J. R. Wijbrans‡, P. M. Shannon§ & T. Morrissey§

* Geomar - Research Center for Marine Geosciences, Wischhofstrasse 1-3, D-24148 Kiel, Germany

† Institute for Geosciences, Christian-Albrechts University, D-24118 Kiel, Germany

‡ Department of Isotope Geochemistry, Vrije University, 1081 HV Amsterdam, The Netherlands

§ Department of Geology, University College Dublin, Belfield, Dublin 4, Ireland

The North Atlantic volcanic province has been attributed to continental rifting about 60 Myr ago over an Iceland plume head with a diameter of 1,000–2,000 km (refs 1, 2). But evidence from a few igneous centres^{3,4} has been used to infer that earlier plume activity occurred in this region. The three seamounts in the Rockall trough off the Atlantic coast of Scotland are among the few accessible remnants of such early plume activity. Here we present ^{40}Ar – ^{39}Ar incremental-heating ages of samples from these seamounts, which show that volcanism began there in the late Cretaceous period (70 ± 1 Myr ago), and then continued for the next 30 Myr in at least four discrete phases: 62, 52, 47 and 42 Myr ago. We relate this activity to pulsing of large masses ($\sim 10^8 \text{ km}^3$) of hot Iceland plume material on timescales of 5–10 Myr. This

significantly extends the time span for Iceland plume activity both backwards and forwards in time, and provides a possible alternative to the 'plume head' models for the formation of continental flood basalts.

In order to establish the origin of the seamounts in the Rockall trough (Fig. 1), we dredged rock samples from their flanks in summer 1998 (Fig. 2; Table 1). Four discrete volcanic episodes (that is, 70 ± 1 Myr, 62 ± 1 Myr, 47 ± 1 Myr and 41 ± 1 Myr ago) were identified in rock samples dredged at five different localities around Anton Dohrn seamount (Table 2; Supplementary Information). This is, to our knowledge, the first reliable rock age evidence for late Cretaceous magmatism in the North Atlantic volcanic province, in agreement with fossil ages for chalks from the eastern flank of Anton Dohrn seamount^{3,4}, and from the top of Rosemary bank seamount³². Ages for rock samples dredged from the seamount on the Hebrides terrace (Table 2; Supplementary Information) indicate three volcanic phases; 62 ± 1 Myr, 51 ± 1 Myr and 48 ± 1 Myr ago. The first and third of these episodes occurred also on Anton Dohrn seamount, while the second event was again sampled on the northwestern flank of Rosemary bank seamount. A 42 ± 1 Myr sample dredged previously⁵ from the southwestern flank of Rosemary bank (Table 2; Supplementary Information) is coeval with the youngest lavas sampled on Anton Dohrn seamount. The weighted average ages for the five discrete Rockall seamount volcanic episodes are 70 ± 1 Myr, 62 ± 1 Myr, 52 ± 1 Myr, 47 ± 1 Myr and 42 ± 1 Myr. All measured ages fall into one of these five distinct volcanic phases, each defining a short, ~ 1 -Myr, period of magmatism, with apparent intervals of ~ 10 Myr separating the first three phases, decreasing to ~ 5 Myr for subsequent ones.

The Rockall seamount samples used in this study are transitional-to-alkali basalts (~ 5 – 8% MgO), characterized by moderately enriched rare-earth-element patterns (Supplementary Information) similar in range to the transitional-to-alkali lavas found on Iceland and other parts of the North Atlantic volcanic province^{6,7}. We see no evidence of marked crustal contamination,

nor for obvious temporal or spatial trends in the geochemical data. Variations in Lu/Hf and Zr/Y ratios indicate that the extent of partial melting was both low and variable⁸. Apparent mixing trends indicated by other incompatible-element ratios sensitive to mixing and source composition (for example, La/Sm, Hf/Th and Tb/Th) confirm the need for at least two distinct end-members. Mixing trends for these transitional-to-alkaline rocks can, as in the case of Iceland, be explained by a thick lithospheric cap restricting melting to small degrees at depth (>65 km), resulting in the more fusible enriched component(s) in the Iceland plume source dominating the depleted component(s) in the melt (see, for example, refs 6–8).

Recent ^{40}Ar – ^{39}Ar (as well as U–Pb and Rb–Sr) dating^{6,9–15}, combined with geochemical information, indicates that volcanic activity related to the Iceland plume began simultaneously in Greenland and along the northwestern European continental margin 60–61 Myr ago. This has been explained by the arrival of a plume head (that is, a 'starting plume')^{6,12,14–16}, rifting above a more slowly growing (that is, 'incubating') plume-fed 'mushroom head'^{11,2}, and by a single event of increased plume flux (that is, a pulse or surge)^{6,14,17,18}, possibly following weaker Cretaceous plume activity⁶. Given that this 60–61-Myr magmatic event also occurred on the Rockall seamounts, together with their being geochemically similar to plume magmatism found on Iceland, we link episodic volcanism on the Rockall seamounts to the activity of the Iceland plume. Both the oldest and youngest ages (60–63 and 52 Myr) measured for the eastern margins of the North Atlantic volcanic province are from the isle of Eigg^{9,10,19}. The apparent absence of magmatism on Eigg during the intervening period¹⁹ tends to support the 62 ± 1 Myr and 52 ± 1 Myr episodicity on the Rockall seamounts. However, many other ^{40}Ar – ^{39}Ar and Rb–Sr ages determined for this region during the past ~ 20 to 30 years straddle this interval, suggesting broadly defined peaks in magmatic activity ~ 59 – 61 Myr, ~ 56 Myr and ~ 52 Myr ago (refs 6, 9, 19, 20). In contrast, many recent incremental-heating ^{40}Ar – ^{39}Ar ages show that plume-related magmatism on Greenland developed as distinct

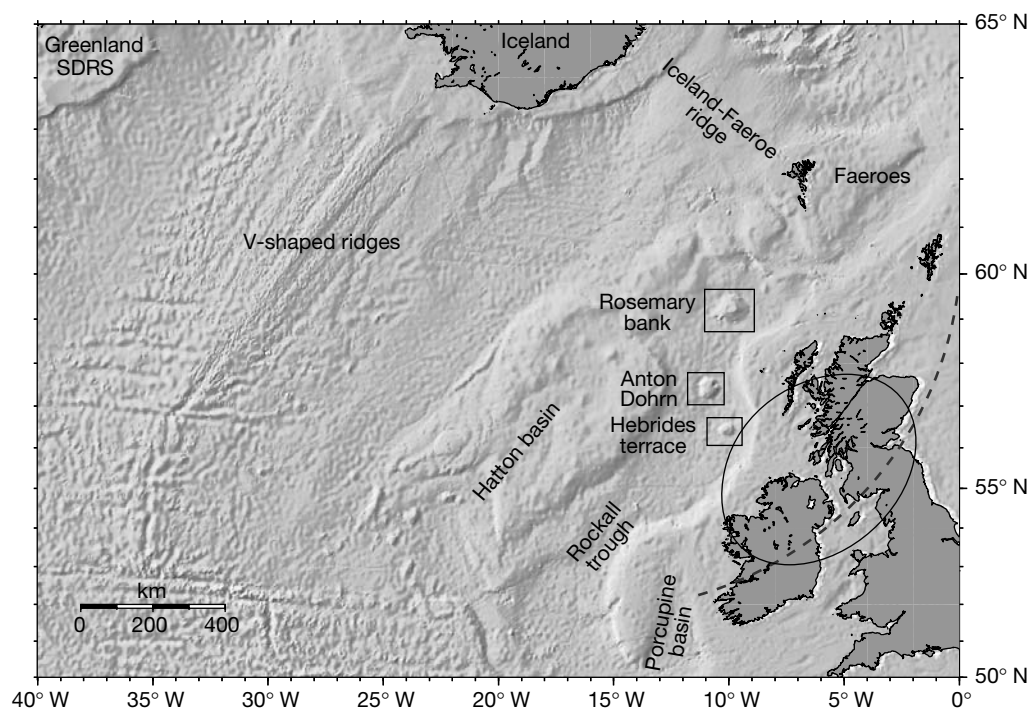


Figure 1 Predicted topography map of the North Atlantic region. The figure shows the location of the Rockall trough seamounts, Iceland (that is, the present location of the plume stem) along with V-shaped ridges to the south. The oval encompasses most onshore Cenozoic continental basalt flows (adapted from ref. 18). The dashed line

indicates the outer reaches of Iceland plume material (inferred from known Iceland plume magmatism) during Rockall seamount development. SDRS, seaward dipping reflector sequence.

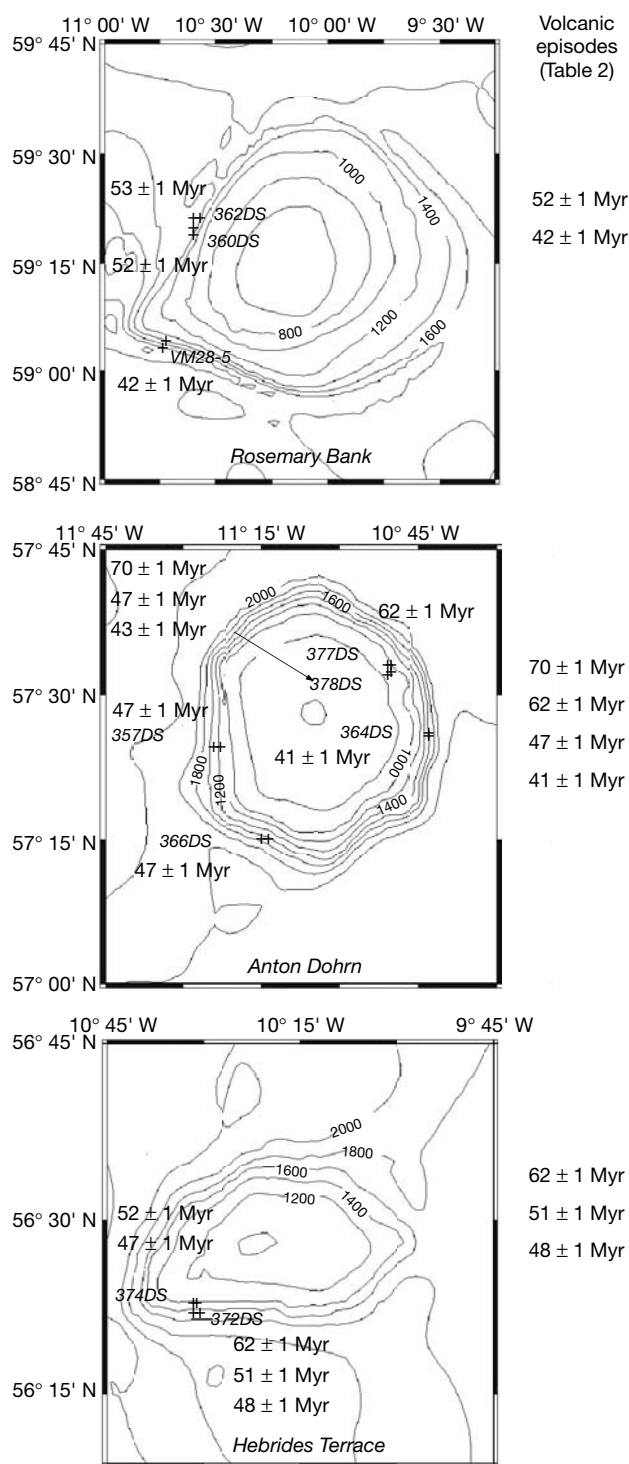


Figure 2 Locations and ^{40}Ar - ^{39}Ar ages of Rockall trough dredge samples. The + symbols indicate start (deeper) and end (shallower) points of dredge tracks during POS 241 sampling stations (Table 1). The weighted average ages (Supplementary Information) for coeval samples are shown beside their respective dredge stations, with the weighted average ages for discrete volcanic episodes sampled at each seamount (Table 2) given beside their respective contour maps (ref. 31; <http://www.bodc.ac.uk/projects/gebco/>). As expected, rocks were also recovered during the POS 241 sampling programme that had probably been transported to the seamounts by glacial drift. However, only samples that were, or had been at some time, *in situ* (for example, evidence of MnO_x coating, pillow or lava flow form, freshly broken surface; details in Supplementary Information) were used in our study. *In situ* origin of selected samples has subsequently been confirmed by geochemical analyses linking them to a common Iceland plume source. Contours are in metres.

episodes^{12–15,21}. In the case of the southwest Greenland seaward-dipping reflector sequence, measured ages for drill samples fall into three distinct time intervals; ~60–61 Myr, ~54–57 Myr and ~49–50 Myr ago (refs 12, 15). Ages for east Greenland continental exposures also define distinct intervals at 55–57 Myr, 47–50 Myr (refs 13, 21) and 33 Myr ago (refs 6, 21), while those for west Greenland show distinct magmatic episodes at ~61 Myr, ~52–55 Myr and ~27–34 Myr ago (ref. 14). With the exception of the apparent ~55–57 Myr interval, magmatic periodicity on Greenland agrees well with that on the Rockall seamounts. Further support for the widespread occurrence of younger Rockall seamount volcanic episodes can be inferred from the prominent V-shaped ridges on the Reykjanes ridge, south of Iceland^{22,23} (Fig. 1). These ridges record the passage of hotter asthenosphere from the centre of the Iceland mantle plume at intervals of 5–10 Myr (refs 22–24). At least six such pulses are indicated for the past ~30 Myr (ref. 24). A pulse or blob of Iceland plume material interacting with the local spreading axis has also been proposed to explain cyclicity in Pb composition on Iceland during the past ~15 Myr (ref. 25).

Episodic magmatism in the Greenland region has so far been explained by discrete magmatic events (or pulses) triggered progressively by plume arrival (~60–61 Myr ago) under central Greenland (that is, starting plume¹⁶), continental breakup (~54–57 Myr ago), and the later passage of the plume axis beneath the east Greenland rifted margin after breakup (~50–49 Myr ago)^{12,13,15}. Furthermore, the younger Eocene ages in west Greenland have been linked to a change in regional plate kinematics and stranded remnants of the Iceland plume head¹⁴. Ultrafast ascent rates for hot mantle plume material (~1–10 myr^{-1}), which spreads out horizontally (~0.5 myr^{-1}) as a thin layer under the lithosphere, have been modelled assuming a non-newtonian (that is, nonlinear) temperature- and depth-dependent upper mantle rheology operating at an effective Rayleigh number of ~10⁶ (refs 17, 26). This provides an alternative to the ‘starting plume’¹⁶ model for North Atlantic flood basalt development by proposing a narrow, but very fast, plume in the upper mantle as a source for the near simultaneous onset of magmatism across the North Atlantic volcanic province 61–60 Myr ago^{17,18}.

Table 1 Sample information

Dredge site	Seamount	Location*	Depth* (m)
POS 241 372	Hebrides Terrace	56° 21.800' N; 10° 30.300' W 56° 23.040' N; 10° 30.620' W	1,659 1,200
POS 241 374	Hebrides Terrace	56° 21.735' N; 10° 31.376' W 56° 22.858' N; 10° 31.561' W	1,698 1,240
POS 241 357	Anton Dohrn	57° 24.722' N; 11° 23.525' W 57° 24.441' N; 11° 22.348' W	1,409 875
POS 241 364	Anton Dohrn	57° 25.361' N; 10° 43.105' W 57° 25.770' N; 10° 42.926' W	1,214 1,035
POS 241 366	Anton Dohrn	57° 14.901' N; 11° 14.799' W 57° 15.300' N; 11° 12.725' W	1,480 945
POS 241 377	Anton Dohrn	57° 33.133' N; 10° 49.415' W 57° 33.019' N; 10° 50.347' W	1,320 956
POS 241 378	Anton Dohrn	57° 32.590' N; 10° 49.020' W 57° 32.310' N; 10° 49.449' W	1,163 949
POS 241 360	Rosemary bank	59° 19.836' N; 10° 35.349' W 59° 19.176' N; 10° 35.254' W	1,478 1,021
POS 241 362	Rosemary bank	59° 20.991' N; 10° 34.725' W 59° 20.928' N; 10° 33.705' W	1,325 1,117
† VM 28-5 P452	Rosemary bank	59° 04.000' N; 10° 42.000' W	2,160 1,830

Detailed sample descriptions available; see Supplementary Information.

* For each dredge site, two locations and depths are given: top entry, on-bottom; lower entry, off-bottom.

† Sample supplied by Lamont Doherty Earth Observatory (ref. 5)

However, neither this, nor any of the previously discussed models, can account for the new information provided here extending the activity of the Iceland plume backwards into the Late Cretaceous, nor the simultaneous development of magmatism related to Iceland plume activity on 5–10 Myr timescales at widely separated locations. Even greater difficulties are presented for these models if the argument is accepted that the V-shaped ridges south of Iceland reflect a continuation of this process for the past ~30 Myr (refs 23, 24). Instead, we suggest that the temporal variations in plume productivity in the North Atlantic volcanic province are related to pulsing of large volumes of hot Iceland plume material, since at least the Late Cretaceous, on timescales of 5–10 Myr. This is compatible with models of ultrafast mantle plumes that show evidence of very focused plumes with calm periods in between, that is, pulsing behaviour ranging in timescale from a few Myr (ref. 17) to about 10 Myr (ref. 26). The most recent such model¹⁷ proposes a pulse rapidly emplacing a ~100 km deep by ~850 km wide (~60 × 10⁶ km³) region of hot plume matter under the lithosphere. This large volume of plume material then migrates rapidly (at 0.5 myr⁻¹) away from the Iceland plume stem region (shown by tomographic studies to be at present rooted under Iceland²⁷). We suggest that magmatism on the Rockall seamounts is related to the Rockall trough being a 'weak or thin spot' in the lithosphere (see, for example, refs 6, 18), which facilitated decompression melting of hot plume material associated with the 'front' of each pulse. This agrees with the apparent relationship between variability in Rockall seamount composition and low, but variable, degrees of partial melting of, presumably, heterogeneous Iceland plume material at >65 km depth (Supplementary Information). Variations in the low degrees of melting can be expected as a result of

changing thickness of the lithosphere under the Rockall trough associated with variability in the temperature of the various plume fronts. No significant change in the composition of the Iceland plume material is therefore, necessarily, required.

The present information about the Iceland plume begs the question as to how temporally representative flood basalts are of plume activity. Pre-basalt doming, extension and alkalic volcanism in other major flood basalt provinces associated with continental breakup (that is, Siberian traps, Karoo province, Parana basin, Rajmahal-Kerguelen and Deccan-Seychelles provinces) suggests that plume 'residence' before the main basalt phase is considerably longer than proposed for the 'starting plume' model (see, for example, ref. 28). The apparent absence on the Rockall seamounts of the 55–57-Myr magmatism observed on Greenland can be explained in terms of the 'incubation' model^{2,28}. Continental rifting following a period of 'incubation' linked to one (or more) plume pulses would have led to large volumes of melt being emplaced only along those borders of the then-diverging continental margins that continued on to full seafloor spreading, in contrast to the Rockall trough, a failed rift. The age of this extensive magmatic event could therefore be out of synchronization with episodic magmatism on the Rockall seamounts, which is linked more directly to the activity of the Iceland plume.

Our results indicate that other mantle plumes may also be pulsing. If so, then their typical pulsing frequencies should lead to a better understanding of upper-mantle circulation; for example, viscosity structure and Rayleigh number. For instance, apparent episodic volcanism on timescales of ~1–2 Myr for the Hawaiian plume trail (see, for example, ref. 29), and ~1 Myr for the Foundation plume³⁰ trail, indicate variable and faster plume ascent

Table 2 Summary of ages for individual seamounts

Seamount	Sample	Plateau (Myr)	±2σ	Inverse isochron (Myr)	±2σ
Hebrides terrace	POS 241 372 DS-2	61.6	0.8	61.6	0.9
Hebrides terrace	POS 241 372 DS-1	50.8	0.6	50.9	0.8
Hebrides terrace	POS 241 374 DS-4	51.9	0.8	52.0	1.2
	Weighted average	51.2	0.5	51.2	0.7
Hebrides terrace	POS 241 372 DS-4	48.2	0.5	47.8	2.1
Hebrides terrace	POS 241 374 DS-1	48.7	0.5	46.8	1.6
	Weighted average	48.5	0.4	47.2	1.3
Anton Dohrn	POS 241 378 DS-2	69.0	2.7	69.1	12.2
Anton Dohrn	POS 241 378 DS-5	70.5	0.9	68.3	15.9
	Weighted average	70.4	0.9	68.8	9.7
Anton Dohrn	POS 241 377 DS-8	62.3	0.6	62.1	0.7
Anton Dohrn	POS 241 377 DS-6	60.8	2.4	55.1	7.1
	Weighted average	62.2	0.6	62.0	0.7
Anton Dohrn	POS 241 357 DS-3	47.7	0.6	47.4	0.7
Anton Dohrn	POS 241 366 DS-1	48.6	2.0	41.6	6.9
Anton Dohrn	POS 241 366 DS-2	47.4	0.8	46.8	0.9
Anton Dohrn	POS 241 366 DS-3	48.5	1.7	48.7	2.3
Anton Dohrn	POS 241 378 DS-3	47.7	0.6	47.2	0.7
Anton Dohrn	POS 241 378 DS-7	47.5	0.6	47.4	0.6
	Weighted average	47.7	0.3	47.3	0.3
Anton Dohrn	POS 241 364 DS-3	41.8	0.9	40.9	1.0
Anton Dohrn	POS 241 364 DS-3	41.3	0.6	40.8	0.9
Anton Dohrn	POS 241 364 DS-6	40.2	2.2	37.3	2.9
Anton Dohrn	POS 241 378 DS-1	42.7	0.5	42.5	1.0
	Weighted average	42.0	0.3	41.2	0.5
Rosemary bank	POS 241 360 DS-3	52.3	0.7	52.2	1.1
Rosemary bank	POS 241 362 DS-1D	53.4	0.9	52.3	1.1
Rosemary bank	POS 241 362 DS-10	53.0	1.5	53.4	1.7
	Weighted average	52.7	0.5	52.4	0.7
Rosemary bank	VM 28-5 P452	42.5	0.4	42.4	0.5

Analytical data, calculated ages and summary of weighted average ages for individual dredge sites, decay and reactor interference constants, as well as isochron and plateau plots, are available; see Supplementary Information. Ages are calculated relative to sanidine monitor TCR (28.34 Myr). Plateau ages are presented as weighted means of concordant steps contributing to the plateau (ranging between 3 and 12 steps); see, for example, ref. 30. Isochron ages can be calculated for the concordant steps in each plateau. In the case of all the analyses reported here, isochron and plateau ages are concordant. ⁴⁰Ar–³⁹Ar incremental-heating analyses were performed by laser probe at Vrije University, Amsterdam. All successful analyses (with the exception of one plagioclase separate) were on whole-rock chips (250–500 μm). Sample preparation and analytical methods have been described in detail elsewhere³⁰. In order to target age variations in the POS 241 dredge samples for incremental-heating experiments, we initially screened a much larger number of POS 241 samples by single fusion ⁴⁰Ar–³⁹Ar dating.

speeds due to greater viscosity differences in an upper mantle operating at a reasonable Rayleigh number ($\sim 10^6$; ref. 17). □

Received 14 February; accepted 10 October 2000.

- White, R. S. *et al.* Magmatism at rifted continental margins. *Nature* **330**, 439–444 (1987).
- White, R. S. & McKenzie, D. P. Magmatism at rift zones: the generation of volcanic continental margins and flood basalts. *J. Geophys. Res.* **94**, 7685–7729 (1989).
- Jones, E. J. W., Ramsay, A. T. S., Preston, N. J. & Smith, A. C. S. A Cretaceous geyser in the Rockall Trough. *Nature* **251**, 129–131 (1974).
- Jones, E. J. W., Siddall, R., Thirwall, M. F., Chronston, P. N. & Lloyd, A. J. Anton Dohrn Seamount and the evolution of the Rockall Trough. *Oceanol. Acta* **17**, 237–247 (1994).
- Dietrich, R. V. & Jones, E. J. W. Volcanic rocks from Rosemary Bank (Rockall Trough, NE Atlantic). *Mar. Geol.* **35**, 287–297 (1980).
- Saunders, A. D., Fittin, J. G., Kerr, A. C., Norry, M. J. & Kent, R. W. in *Large Igneous Provinces: Continental, Oceanic, and Planetary Flood Volcanism* (eds Mahoney, J. J. & Coffin, M. F.) 45–93 (Geophysical Monograph 100, American Geophysical Union, Washington DC, 1997).
- Kerr, A. C. The melting processes and composition of the North Atlantic (Iceland) plume: geochemical evidence from the Early Tertiary basalts. *J. Geol. Soc. Lond.* **152**, 975–978 (1995).
- Fram, M. S., Leshner, C. E. & Volpe, A. M. Mantle melting systematics: Transition from continental to oceanic volcanism on the southeast Greenland margin. *Proc. ODP Sci. Res.* **152**, 373–386 (1998).
- Mussett, A. E. & McCormack, A. G. *Magnetic Polarity Timescales: a New Test 27–37* (Special Publication 70, Geological Society, London, 1992).
- Pearson, D. G., Emeleus, C. H. & Kelley, S. P. Precise $^{40}\text{Ar}/^{39}\text{Ar}$ age for the initiation of Palaeogene volcanism in the Inner Hebrides and its regional significance. *J. Geol. Soc. Lond.* **153**, 815–818 (1996).
- Hamilton, M. A., Pearson, D. G., Thompson, R. N., Kelley, S. P. & Emeleus, C. H. Rapid eruption of Skye lavas inferred from precise U–Pb and Ar–Ar dating of the Rum and Cuillin plutonic complexes. *Nature* **394**, 260–263 (1998).
- Sinton, C. W. & Duncan, R. A. $^{40}\text{Ar}/^{39}\text{Ar}$ ages of lavas from the southeast Greenland margin, ODP Leg 152, and the Rockall Plateau, DSDP Leg 81. *Proc. ODP Sci. Res.* **152**, 387–402 (1998).
- Tegner, C. *et al.* $^{40}\text{Ar}/^{39}\text{Ar}$ geochronology of Tertiary mafic intrusions along the East Greenland rifted margin: Relation to flood basalts and the Iceland hotspot track. *Earth Planet. Sci. Lett.* **156**, 75–88 (1998).
- Storey, M., Duncan, R. A., Pedersen, A. K., Larsen, L. M. & Larsen, H. C. $^{40}\text{Ar}/^{39}\text{Ar}$ geochronology of the West Greenland Tertiary volcanic province. *Earth Planet. Sci. Lett.* **160**, 569–586 (1998).
- Tegner, C. & Duncan, R. A. $^{40}\text{Ar}/^{39}\text{Ar}$ chronology for the volcanic history of the southeast Greenland rifted margin. *Proc. ODP Sci. Res.* **163**, 53–62 (1999).
- Richards, M. A., Duncan, R. A. & Courtillot, V. E. Flood basalts and hot-spot: Plume heads and tails. *Science* **246**, 103–107 (1989).
- Larsen, T. B., Yuen, D. A. & Storey, M. Ultrafast mantle plumes and implications for flood basalt volcanism in the Northern Atlantic Region. *Tectonophysics* **311**, 31–43 (1999).
- Larsen, H. C. & Saunders, A. D. Tectonism and volcanism at the southeast Greenland rifted margin: A record of plume impact and later continental rupture. *Proc. ODP Sci. Res.* **152**, 503–533 (1998).
- Mussett, A. E., Dagley, P. & Skelhorn, R. R. *Time and Duration of Activity in the British Tertiary Igneous Province 337–348* (Special Publication 39, Geological Society, London, 1988).
- Meighan, I. G., McCormick, A. G., Gibson, D., Gamble, J. A. & Graham, I. J. *Rb–Sr Isotopic Determinations and the Timing of Tertiary Central Complex Magmatism in NE Ireland 349–360* (Special Publication 39, Geological Society, London, 1988).
- Upton, B. G. J., Emeleus, C. H., Rex, D. C. & Thirwall, M. F. Early Tertiary magmatism in NE Greenland. *J. Geol. Soc. Lond.* **152**, 959–964 (1995).
- Vogt, P. R. Asthenosphere motion recorded by the ocean floor south of Iceland. *Earth Planet. Sci. Lett.* **13**, 153–160 (1971).
- White, R. S., Brown, J. W. & Smallwood, J. R. The temperature of the Iceland plume and the origin of outward-propagating V-shaped ridges. *J. Geol. Soc. Lond.* **152**, 1039–1045 (1995).
- White, N. & Lovell, B. Measuring the pulse of a plume with the sedimentary record. *Nature* **387**, 888–891 (1997).
- Hanan, B. B. & Schilling, J.-G. The dynamic evolution of the Iceland mantle plume: the lead isotope perspective. *Earth Planet. Sci. Lett.* **151**, 43–60 (1997).
- Larsen, T. B. & Yuen, D. A. Ultrafast upwelling bursting through the upper mantle. *Earth Planet. Sci. Lett.* **146**, 393–399 (1997).
- Bijwaard, H. & Spakman, W. Tomographic evidence for a narrow whole mantle plume below Iceland. *Earth Planet. Sci. Lett.* **166**, 121–126 (1998).
- Kent, R. W., Storey, M. & Saunders, A. D. Large igneous provinces: Sites of plume impact or plume incubation? *Geology* **20**, 891–894 (1992).
- Clague, D. A. & Dalrymple, G. B. in *The Eastern Pacific Ocean and Hawaii: The Geology of North America* (eds Winterer, E. L., Hussong, D. M. & Decker, R. W.) 187–286 (Geological Society of America, Boulder, Colorado, 1989).
- O'Connor, J. M., Stoffers, P. & Wijbrans, J. R. Migration rate of volcanism along the Foundation Chain, SE Pacific. *Earth Planet. Sci. Lett.* **164**, 41–59 (1998).
- GEBCO Digital Atlas* (British Oceanographic Data Centre, Bidston, 1997).
- Morton, A. C. *et al.* Late Cretaceous basalts from Rosemary Bank, Northern Rockall Trough. *J. Geol. Soc. Lond.* **152**, 947–952 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank Captain Gross and the crew of FS *Poseidon*, along with the scientific party, for ship handling and support during sampling, and T. Arpe and D. Garbe-Schoenberg for ICP-MS data. LDEO supplied a Rosemary bank rock sample. J.O.C. was affiliated to AWI-Bremerhaven during much of this study. This work was supported by the DFG.

Correspondence and requests for information should be addressed to J.O.C. (e-mail: joconnor@geomar.de).

Subduction and collision processes in the Central Andes constrained by converted seismic phases

X. Yuan^{*†}, S. V. Sobolev^{*}, R. Kind^{*†}, O. Oncken^{*†}, G. Bock^{*||}, G. Asch^{*||}, B. Schurr^{*||}, F. Graeber^{*||}, A. Rudloff^{*||}, W. Hanka^{*||}, K. Wylegalla^{*||}, R. Tibi^{*||}, Ch. Haberland^{†||}, A. Rietbrock^{†||}, P. Giese^{†||}, P. Wigger^{†||}, P. Röwer^{†||}, G. Zandt^{†||}, S. Beck^{†||}, T. Wallace^{†||}, M. Pardo^{§||} & D. Comte^{§||}

^{*} GeoForschungsZentrum Potsdam, Telegrafenberg, 14473 Potsdam, Germany

[†] Freie Universität Berlin, Fachrichtung Geophysik, Malteser Str. 74–100, 12249 Berlin, Germany

[‡] University of Arizona, Department of Geosciences, Tucson, Arizona 85721, USA

[§] Universidad de Chile, Departamento de Geofísico, Blanco Encalada 2085, Santiago, Chile

^{||} Andes Working Group

The Central Andes are the Earth's highest mountain belt formed by ocean–continent collision^{1,2}. Most of this uplift is thought to have occurred in the past 20 Myr, owing mainly to thickening of the continental crust^{2–6}, dominated by tectonic shortening^{7–10}. Here we use P-to-S (compressional-to-shear) converted teleseismic waves observed on several temporary networks in the Central Andes to image the deep structure associated with these tectonic processes. We find that the Moho (the Mohorovičić discontinuity—generally thought to separate crust from mantle) ranges from a depth of 75 km under the Altiplano plateau to 50 km beneath the 4-km-high Puna plateau. This relatively thin crust below such a high-elevation region indicates that thinning of the lithospheric mantle may have contributed to the uplift of the Puna plateau. We have also imaged the subducted crust of the Nazca oceanic plate down to 120 km depth, where it becomes invisible to converted teleseismic waves, probably owing to completion of the gabbro–eclogite transformation; this is direct evidence for the presence of kinetically delayed metamorphic reactions in subducting plates. Most of the intermediate-depth seismicity in the

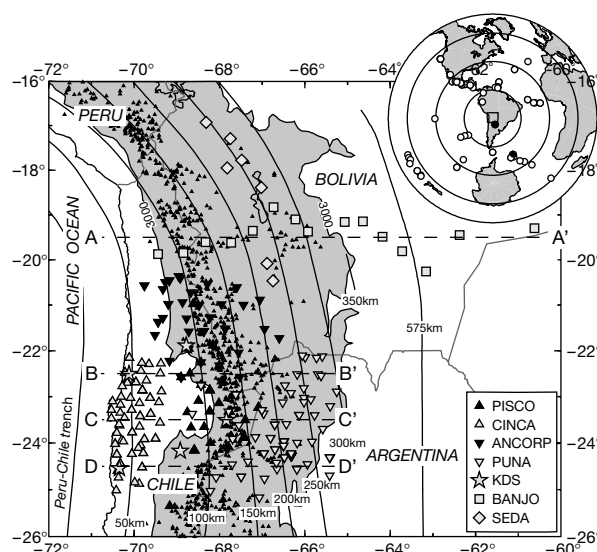


Figure 1 Map of the Central Andes, showing networks of passive seismological stations used in this study. Areas with elevation higher than 3 km are shaded. Contour lines mark the Wadati–Benioff zone²⁷. Small triangles mark volcanoes. Inset, locations of several teleseismic earthquakes and two deep regional earthquakes used in this study. Data from two earthquakes marked by filled circles show strong P-to-S conversions of the subducted Nazca plate, and are modelled in Fig. 4.

subducting plate stops at 120 km depth as well, suggesting a relation with this transformation. We see an intracrustal low-velocity zone, 10–20 km thick, below the entire Altiplano and Puna plateaux, which we interpret as a zone of continuing metamorphism and partial melting that decouples upper-crustal imbrication from lower-crustal thickening.

The Central Andes are a unique place where subduction and mountain-building processes can be studied, both in their extremes. Whereas subduction has been continuing at least since the early Mesozoic era, evolution of the Andean cordillera has been concentrated in the Cenozoic era during continuous subduction of the Nazca plate under the Central Andes at a velocity between 6 and 15 cm yr⁻¹ (refs 1, 2). The Moho of the continental upper plate has been detected by different seismic methods at 30–80 km depth, with its deepest parts below the Eastern and Western Cordilleras, which border the Altiplano and Puna plateaux^{3–6}. The altitude of the

plateaux (about 4 km) has been explained as an isostatic response to crustal thickening^{2,6}. A low average P-wave velocity and low Poisson's ratio in the crust of the northern Altiplano have been reported from seismic observations, indicating tectonic shortening as the dominant mechanism of crustal thickening in the northern Central Andes^{5,6}. Geological studies, however, report only up to 250 km of crustal shortening (equivalent to some 70% of the width of the Altiplano^{7–9}) and only some 75 km of shortening at the latitude of Puna¹⁰. Our aim was to resolve crustal and mantle structures associated with the plateau building and subduction processes in the Central Andes.

Several seismic networks were operated between 1994 and 1997 in northern Chile, southern Bolivia and northwestern Argentina (Fig. 1) for periods ranging from two months to one year. Receiver functions (RFs) (P-to-S converted waves isolated from the incident teleseismic P waves) are commonly used to constrain crustal velocities and to detect crustal and upper-mantle discontinuities underneath seismic stations^{11,12}. Figure 2a shows the first 15 s after P onset of about 500 averaged RFs displayed on an east–west profile between 65° and 71° W. Besides the strong phases marked B (at less than 1 s) and M (at 6–8 s), which are conversions from the basement and the Moho, there are some other pronounced intracrustal phases that can be traced across the entire Altiplano and Puna. They are called trans-Andean converters (TRAC) and in Fig. 2 are marked TRAC1 (blue, negative, velocity decrease downwards) and TRAC2 (red, positive, velocity increase downwards). The subducting Nazca plate is also visible.

As a dense seismic network is available, it is helpful to migrate the RFs into the space domain to image seismic discontinuities^{13,14}—a process similar to the near-vertical incidence reflection method in crustal studies. Figure 2c shows, as background to an interpretative cartoon, the time domain RFs of Fig. 2a migrated into a depth domain image of the crust and upper mantle of the Central Andes down to 200 km depth. The most pronounced feature in Fig. 2a and c is the continental Moho converter at 40–75 km depth. It is consistent with the previous estimates of crustal thickness in the Central Andes from seismological observations^{5,6} and from wide-angle reflections⁴. Images of the Moho along four lines crossing the Central Andes (see Fig. 1 for location of lines) are presented in Fig. 3

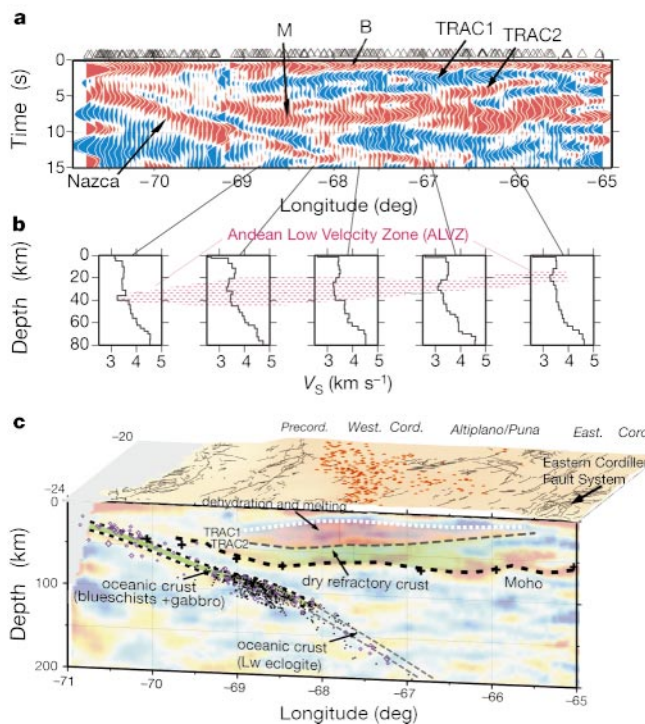


Figure 2 Receiver function (RF) images and crustal models of the Central Andes along an east–west profile. In the images, red indicates positive, velocity increase downwards; blue indicates negative, velocity reduction downwards. **a**, Time domain RFs averaged over a 30-km-wide moving window, 2–10-s band pass filtered. **b**, Possible crustal S-wave velocity (v_s) models resulting from the inversion of the RFs in **a**. **c**, Interpretative cartoon with depth-migrated RF image as background averaged over the Altiplano and Puna regions. Black plus signs show Moho depth data from wide-angle reflection studies⁴. Local crustal models and a global mantle reference model were used for migration. The TRAC1 and TRAC2 converters in **a** and **c** border the Andean low-velocity zone (ALVZ) which dips westwards from below the Eastern Cordillera fault system (indicated in **c** by black lines at the surface) across the entire Altiplano/Puna to the Precordillera (red points at surface indicate Cenozoic volcanoes). Figure S1 in Supplementary Information shows this data set split in two at 23° S for judging north–south variations of the TRAC2 converter. Figure S2 in Supplementary Information shows the lateral extension of the ALVZ. The Nazca converter (thick black-white dashed line indicated in **c**) is interpreted as an image of the oceanic Moho. The upper boundary of the oceanic crust (slab shear zone) is set 10 km above the oceanic Moho in agreement with the results of waveform inversion (see text and Fig. 4). Oceanic crust is clearly visible from converted waves only down to 120 km depth. Hypocentres of the intermediate depth earthquakes are also shown. Black dots are small-magnitude events ($M = 2-4$) located by the PISCO network during 3 months of operation²⁸. Purple diamonds are 1965–95 relocated events²⁹ at 22–24° S, with the depth corrected for the local velocity model in the Central Andes. Only the best located events that have more than 10 depth phases are shown. Their magnitudes, ranging from 4.5 to 6.5, are indicated by the symbol size. Lw, the mineral lawsonite.

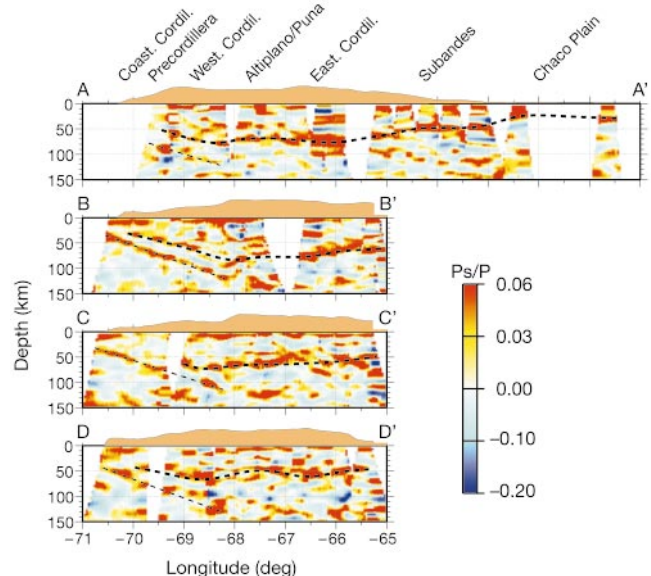


Figure 3 Four east–west depth profiles of migrated RF data within latitude ranges of 1° each. The locations of the profiles are shown in Fig. 1. The surface topography is averaged also within latitude ranges of 1° each. Dashed lines mark the continental and the subducted oceanic Mohos. We note significant uplift of the Moho from the Altiplano (lines A–A' and B–B') to the Puna (lines C–C' and D–D') at 66–68° W, which is not accompanied by large altitude changes.

together with the smoothed surface topography. Large variations of the depth of the Moho beneath the high plateau indicate strong heterogeneity of the lithospheric thickness, where a thin mantle lithosphere is required to maintain high topography if the crust is thin. This is probably the case for the Puna plateau, in accord with strong lithospheric delamination suggested for this region¹⁵.

The second pronounced feature is the Nazca converter (Fig. 2c), which can be traced parallel to the Wadati–Benioff zone down to a depth of 120 km. Individual P-to-S conversions from the Nazca converter are most clearly seen in records from two earthquakes with relatively flat incidence angles on the subducting plate, in agreement with theoretical expectations¹⁶ (see Fig. 4 and filled circles in Fig. 1 for epicentres). A dominant dipping positive phase (marked black in Fig. 4) crosses the entire CINCA network from 5.5 s to 9.5 s from west to east and continues further eastwards into the centre of the PISCO network at 14.5 s delay time. (The locations of the CINCA and PISCO networks are shown in Fig. 1.) About 1.3 s earlier than the positive phase, there is a clear correlatable negative phase, which cannot be interpreted as a side lobe of the P signal because it is too large and appears only on one side of the positive phase.

The RF waveforms can be modelled by a low-velocity layer, 5–10 km thick, with an S-wave velocity contrast of about 15% (inset in Fig. 4). This layer thickness and velocity contrast to the surrounding mantle corresponds to non-eclogitized mafic oceanic crust versus garnet peridotite¹⁷. In this case, the Nazca converter represents the Moho of the subducted oceanic crust. As seen from Fig. 2c, the Nazca converter is exceptionally ‘bright’ down to a depth of about 120 km, where it terminates abruptly. The probable cause of such a breakdown is completion of the gabbro to eclogite transformation in most of the oceanic crust, which would make it seismically almost

indistinguishable from the peridotitic mantle¹⁷. As eclogite is the thermodynamically stable modification of both water-containing and dry mafic rocks at depths greater than 80 km (ref. 18), observation of non-eclogitized crust deeper than 80 km provides direct evidence for kinetically delayed metamorphic reactions in the subducting plate—such reactions have been invoked previously to explain intermediate depth earthquakes¹⁹.

The migrated RF image (Fig. 2c) of the subducting crust addresses directly the question of whether intermediate earthquakes are located within the crust or not. High-precision earthquake locations in the Central Andes between 22° and 24° S are also shown in Fig. 2c. Most of the events are located within the oceanic crust which is still not completely eclogitized, but the lower part of the middle section of the cluster extends below the oceanic crust by some 10–20 km. This observation differs from the situation at 21° S, where most of the cluster is apparently shifted into the oceanic mantle, as reported by a deep reflection study²⁰. Most of the intermediate depth seismicity, as well as the RF image of incompletely eclogitized crust, terminates where a conspicuous increase in subduction angle occurs. This may be an additional indication of the same reaction-driven processes within the slab. In this case enhancement of slab pull following the completion of eclogitization and/or decrease of oceanic crustal strength due to the continuing dehydration may control this feature.

Two clear intracrustal converters TRAC1 and TRAC2 (Fig. 2a–c) confine a low-velocity zone (hereafter called the Andean low-velocity zone, ALVZ). The ALVZ crosses the entire Altiplano/Puna plateau from the Eastern Cordillera to the Western Cordillera, dipping slightly to the west, and is characterized by a reduction of seismic wave velocity of about 10–20% (see Fig. 2b). Locally, this reduction is even higher, with the highest values below major ignimbrite fields (Supplementary Information, Fig. S2). The existence of the extended low-velocity zone was previously reported from wide-angle reflection studies along several lines crossing the Central Andes⁴. From clear data of P-to-S conversions and their multiples (Supplementary Information, Fig. S3), we estimate the average crustal v_p/v_s ratio to be about 1.80 in the southern Altiplano and Puna (here v_p and v_s are the velocities of P and S waves, respectively). This value of 1.80 is much larger than the value of 1.70 that is expected for solid elastic rocks having v_p values less than 6.3 km s^{-1} in the α -quartz stability field¹⁷. This indicates some significant non-lithological effects on seismic velocities, or very high temperatures (β -quartz stability). To the north of 22° S (see Supplementary Information, Fig. S2), the ALVZ is also evident but is weaker, in accord with a lower v_p/v_s ratio observed there^{5,6}.

Due to the clear correlation of the ALVZ strength with volcanic activity (Fig. S2, Supplementary information) and high v_p/v_s ratio, we interpret the ALVZ as a high-temperature zone with partial melting achieved in the places of strongest conversions (Supplementary Information, Fig. S2) and continuing prograde metamorphic reactions in the rest of it. The thermal conditions needed for this feature are most probably caused by mantle magmatism and heat advection (related to subduction, lithospheric mantle delamination¹⁵ and intracrustal diapirism), as thickening of radiogenic crust and shear heating could not have achieved mid-crustal melting since the Miocene epoch. Hot lower crust before shortening, as indicated from xenolith data of crust originally located to the east of the Eastern Cordillera²¹, may have supported this evolution. In this interpretation, the top of the ALVZ (TRAC1 boundary) represents the upper limit of compaction-porous vertical flow of fluids and melts associated with the present brittle–ductile transition in the upper crust. Melt accumulation along this boundary may produce huge partial-melt bodies²² (see also Supplementary Information, Fig. S2). The bottom of the ALVZ (TRAC2 boundary) may represent the upper boundary of the dry and hot refractory lower crust, which has already lost melts and volatiles and therefore obtained higher seismic velocities.

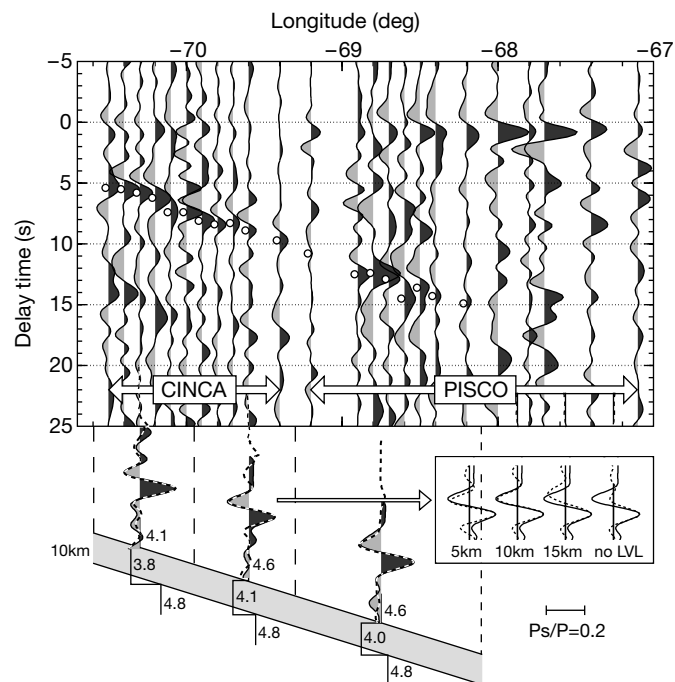


Figure 4 RFs from two earthquakes (filled circles in Fig. 1) recorded by the CINCA and PISCO seismic networks and their modelling. Single RFs are sorted from west to east by the geographical longitude of stations and stacked in bins of 10 km. A 2–10-s bandpass filter is applied to the data. P-to-S conversions can be clearly traced from 5.5 s to 14.5 s time. Positive amplitudes of the converted waves (black) are coherently preceded by negative amplitudes (grey). Stacked RF waveforms can be modelled by the response of a 5–10-km-thick layer with a reduction of S-wave velocity of about 15% relative to the surrounding mantle. A model without a low-velocity layer (LVL) is unable to explain the large negative signals in front of the positive Nazca converter.

Moreover, the direct link of the ALVZ to the basal detachment under the eastern plateau margin, as well as its correlation with the extent of the entire shortened plateau upper crust (Fig. 2), strongly supports its relevance to the tectonic shortening process. The mechanical weakness of the mid-crustal ALVZ and the Eastern Cordillera basal detachment, as inferred from the presence of melts and fluids, probably separates crustal shortening in an upper-crust imbricate belt^{7–9} from a deeper crust with an unknown mode of internal deformation. The continuity of the deep crustal structure across the entire plateau suggests a single crustal thickening process, namely tectonic shortening, rather than important additional contributions from other processes that involve the mantle²³. As a consequence, bulk crustal shortening from surface observations may have been substantially underestimated.

Our interpretation of the ALVZ is similar, in its tectonic aspect, to the interpretation of the low-velocity zone in Tibet—which has remarkably similar depth, thickness and lateral extent^{24,25}. Moreover, as documented by a study of xenoliths²⁶, the middle crust (30–50 km depth) of the northern Tibetan plateau consists of hot and dry refractory rocks similar to those we expect below TRAC2. The existence of analogous low-velocity zones below the Earth's two highest plateaux suggests their fundamental role in the plateau-building process, possibly as decoupling zones that partition crustal shortening into a largely brittle upper-crust domain and ductile deep crust. □

Received 15 October 1999; accepted 12 October 2000.

- DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. Current plate motions. *Geophys. J. Int.* **101**, 425–478 (1990).
- Isacks, B. L. Uplift of the central Andean plateau and bending of the Bolivian orocline. *J. Geophys. Res.* **93**, 3211–3231 (1988).
- James, D. E. Andean crustal and upper mantle structure. *J. Geophys. Res.* **76**, 3246–3271 (1971).
- Wigger, P. J. *et al.* in *Tectonics of the Southern Central Andes* (eds Reutter, K. J., Scheuber, E. & Wigger, P. J.) 23–48 (Springer, Berlin, 1994).
- Zandt, G., Velasco, A. A. & Beck, S. L. Composition and thickness of the southern Altiplano crust, Bolivia. *Geology* **22**, 1003–1006 (1994).
- Beck, S. L. *et al.* Crustal thickness variations in the Central Andes. *Geology* **24**, 407–410 (1996).
- Allmendinger, R. W. & Gubbels, T. Pure and simple shear plateau uplift, Altiplano-Puna, Argentina and Bolivia. *Tectonophysics* **259**, 1–13 (1996).
- Kley, J. & Monaldi, C. R. Tectonic shortening and crustal thickness in the central Andes: how good is the correlation? *Geology* **26**, 723–726 (1998).
- Lamb, S., Hoke, L., Kennan, L. & Dewey, J. in *Orogeny Through Time* (eds Burg, J. P. & Ford, M.) 237–264 (Special Publication 121, Geological Society, London, 1997).
- Allmendinger, R. W., Jordan, T. E., Kay, S. M. & Isacks, B. L. The evolution of the Altiplano-Puna Plateau of the Central Andes. *Annu. Rev. Earth Planet. Sci.* **26**, 139–174 (1997).
- Owens, T. J., Zandt, G. & Taylor, S. R. Seismic evidence for an ancient rift beneath the Cumberland Plateau, Tennessee: A detailed analysis of broadband teleseismic P waveforms. *J. Geophys. Res.* **89**, 7783–7795 (1984).
- Yuan, X., Ni, J., Kind, R., Mechie, J. & Sandvol, E. Lithospheric and upper mantle structure of southern Tibet from a seismological passive source experiment. *J. Geophys. Res.* **102**, 27491–27500 (1997).
- Jones, C. H. & Phinney, R. A. Seismic structure of the lithosphere from teleseismic converted arrivals observed at small arrays in the southern Sierra Nevada and vicinity, California. *J. Geophys. Res.* **103**, 10065–10090 (1998).
- Kosarev, G. *et al.* Seismic evidence for a detached Indian lithospheric mantle beneath Tibet. *Science* **283**, 1306–1309 (1999).
- Kay, R. W. & Kay, S. M. Delamination and delamination magmatism. *Tectonophysics* **219**, 177–189 (1993).
- Cassidy, J. F. Numerical experiments in broadband receiver function analysis. *Bull. Seismol. Soc. Am.* **82**, 1453–1474 (1992).
- Sobolev, S. V. & Babeyko, A. Yu. Modelling of mineralogical composition, density and elastic wave velocities in the unhydrous rocks. *Surv. Geophys.* **15**, 515–544 (1994).
- Peacock, S. M. The importance of blueschist - eclogite dehydration reactions in subducting oceanic crust. *Geol. Soc. Am. Bull.* **105**, 684–694 (1993).
- Kirby, S., Engdahl, E. R. & Denlinger, R. in *Subduction Top to Bottom* (eds Bebout, G. E., Scholl, D. W., Kirby, S. H. & Platt, J. P.) 195–214 (Geophysical Monograph 96, American Geophysical Union, Washington DC, 1996).
- ANCORP Working Group. Seismic reflection image of Andean subduction zone revealing offset of intermediate-depth seismicity into oceanic mantle. *Nature* **397**, 341–344 (1999).
- Lucassen, F., Lewenz, S. & Franz, G. Metamorphism, isotopic ages and composition of lower crustal granulite xenoliths from the Cretaceous Salta Rift, Argentina. *Contrib. Mineral. Petrol.* **134**, 325–341 (1999).
- Chmielewski, J., Zandt, G. & Haberland, C. The central Andean Altiplano-Puna Magma body. *Geophys. Res. Lett.* **26**, 783–786 (1999).
- Giese, P., Scheuber, E., Schilling, F., Schmitz, M. & Wigger, P. Crustal thickening processes in the Central Andes and the different natures of the Moho-discontinuity. *J. S. Am. Earth Sci.* **12**, 201–220 (1999).
- Nelson, K. D. *et al.* Partially molten middle crust beneath southern Tibet: Synthesis of project INDEPTH results. *Science* **274**, 1684–1688 (1996).

- Kind, R. *et al.* Evidence from earthquake data for a partially molten crustal layer in southern Tibet. *Science* **274**, 1692–1694 (1996).
- Hacker, B. R. *et al.* Hot and dry deep crustal xenoliths from Tibet. *Science* **287**, 2463–2466 (2000).
- Cahill, T. A. & Isacks, B. L. Seismicity and shape of the subducted Nazca Plate. *J. Geophys. Res.* **97**, 17503–17529 (1992).
- Graeber, F. & Asch, G. Three dimensional models of P wave velocity and P to S velocity ratio in the southern central Andes by simultaneous inversion of local earthquake data. *J. Geophys. Res.* **104**, 20237–20256 (1999).
- Engdahl, E. R., van der Hilst, R. D. & Buland, R. Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bull. Seismol. Soc. Am.* **88**, 722–743 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank R. Trumbull and F. Lucassen for discussions, J. Mechie and D. Harlov for comments on the manuscript, and G. Chong and M. Wilke for supporting the experiments. The field experiments were supported by the Collaborative Research Center (SFB) 267 of the Deutsche Forschungsgemeinschaft, the GeoForschungsZentrum Potsdam, the Freie Universität Berlin, the US National Science Foundation, the PASSCAL project, the Universidad de Chile (Santiago) and the Universidad Católica del Norte (Antofagasta).

Correspondence and requests for materials should be addressed to R.K. (e-mail: kind@gfz-potsdam.de).

Disturbance and diversity in experimental microcosms

Angus Buckling*, Rees Kassen†, Graham Bell‡ & Paul B. Rainey*

* Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK

† Department of Biology, McGill University, Montreal H3A 1B1, Quebec, Canada

‡ Redpath Museum, McGill University, Montreal H3A 2K6, Quebec, Canada

External agents of mortality (disturbances) occur over a wide range of scales of space and time, and are believed to have large effects on species diversity. The “intermediate disturbance hypothesis”^{1–3}, which proposes maximum diversity at intermediate frequencies of disturbance, has received support from both field^{4,5} and laboratory^{6,7} studies. Coexistence of species at intermediate frequencies of disturbance is thought to require trade-offs between competitive ability and disturbance tolerance⁸, and a metapopulation structure, with disturbance affecting only a few patches at any given time^{9–11}. However, a unimodal relationship can also be generated by global disturbances that affect all patches simultaneously, provided that the environment contains spatial niches to which different species are adapted¹². Here we report the results of tests of this model using both isogenic and diverse populations of the bacterium *Pseudomonas fluorescens*. In both cases, a unimodal relationship between diversity and disturbance frequency was generated in heterogeneous, but not in homogeneous, environments. The cause of this relationship is competition among niche-specialist genotypes, which maintains diversity at intermediate disturbance, but not at high or low disturbance. Our results show that disturbance can modulate the effect of spatial heterogeneity on biological diversity in natural environments.

Selection in heterogeneous environments may act to maintain diversity^{13–15}, and the conditions under which this can occur have recently been derived¹⁶ using a modification of the classic Levene model¹⁷. There are two key requirements. First, each species must be fitter than the other in one of the two niches, so that selection is antagonistic. Second, each niche must contribute approximately equal numbers of individuals to the global population. If one niche contributes many more individuals to the community than the

other, the type that is better adapted to this niche will become fixed, even if it is greatly inferior in the less productive niche. We have previously shown how productivity affects diversity through its effect on the relative production of niches¹⁶. We now show that disturbance can be treated in a similar way.

The frequency of disturbance is likely to affect diversity in a heterogeneous environment because the production of individuals from each niche will be a function of time since the last disturbance. Consider a population growing logistically¹⁸ with an initial time lag during which the individuals adjust physiologically to local conditions but do not reproduce. If disturbance is very frequent, the niche associated with the smaller time lag will produce the most individuals, and at equilibrium the community will be dominated by the type better adapted to this niche (Figs 1a and 2a). Conversely, if disturbance is very infrequent the niche with the greater carrying capacity will contribute more individuals to the global population, and the type adapted to this niche will dominate (Figs 1a and 2a). At intermediate levels of disturbance, the number of individuals emerging from each niche becomes approximately equal, which permits the two types to coexist. This generates a unimodal

relationship between diversity and disturbance frequency (Figs 1b and 2b). Depending on parameter values, this may involve either the steady replacement of one type by another along a disturbance gradient (Fig. 1b) or dominance of the same type at both high and low disturbance frequency (Fig. 2b). We note that it is possible to obtain a similar relationship between disturbance frequency and diversity without including a time lag, although the conditions under which this occurs are more restrictive.

To test the hypothesis that diversity is unimodally related to the frequency of global disturbance in heterogeneous, but not homogeneous, environments we used populations of the bacterium *P. fluorescens*^{19,20}. In a spatially heterogeneous environment (a static glass microcosm containing nutrient broth medium) *P. fluorescens* populations rapidly diversify, generating a range of specialist genotypes that are readily distinguished by their colony morphologies on agar plates¹⁵. Three main classes of specialist are readily distinguishable: smooth (SM) morphs, resembling the ancestral genotype, which inhabit the liquid phase; wrinkly-spreading (WS) morphs, which colonize the air–broth interface; and fuzzy-spreader (FS) morphs, which colonize the bottom of the vial. Within each main category substantial variation exists. Populations evolving in spatially homogeneous environments (continuously shaken microcosms) show little evidence of diversification and are predominantly of the SM type¹⁵. Because reproduction is entirely asexual, these variants are analogous to species¹⁶.

Two experiments differing in the amount of diversity present in the founding populations were conducted simultaneously. In the first, microcosms were founded by populations that had previously

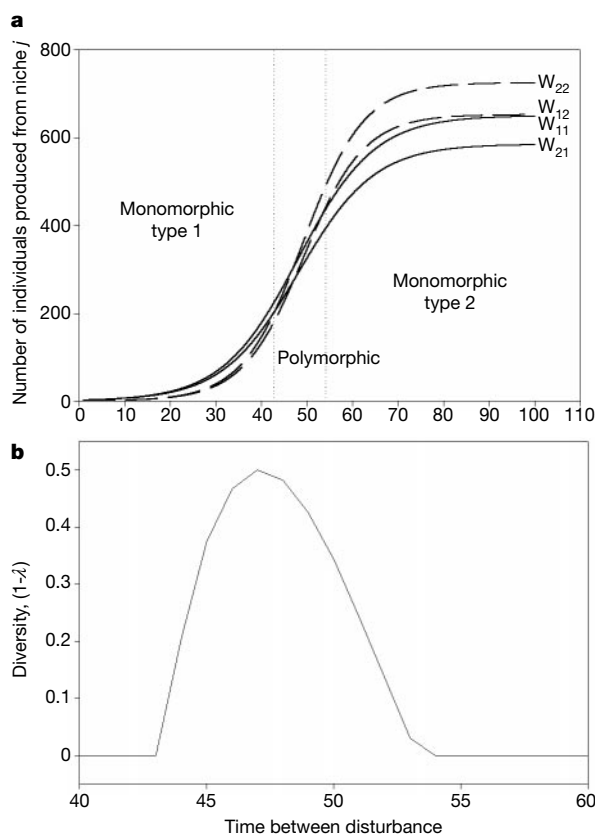


Figure 1 The effect of disturbance frequency on diversity, in the case where competitive ability and disturbance tolerance are negatively related. **a**, The number of individuals of type j produced from niche j as a function of time between disturbance follows the logistic model:

$$N_j(t) = \frac{K_j}{1 + a_j e^{-r_j(t-T_j)}} \quad (1)$$

where K is the carrying capacity of a given niche, a_j is a constant defined as $[(K_j - N_j(0))/N_j(0)]$, where $N(0)$ is density at time 0, r is the limiting rate of growth, t is time, and T is the lag or acclimation period that elapses before growth begins again. Parameter values are: type 1 in niche 1 (W_{11}), $r = 0.12$, $K = 650$, $T = 0$; type 2 in niche 2 (W_{22}), $r = 0.15$, $K = 725$, $T = 10$; $N_0 = 2$ in all cases. The less fit type in each niche had a fitness equal to 90% of that of the more fit type at all time points. **b**, The behaviour of the diversity index, $1 - \lambda$, at equilibrium.

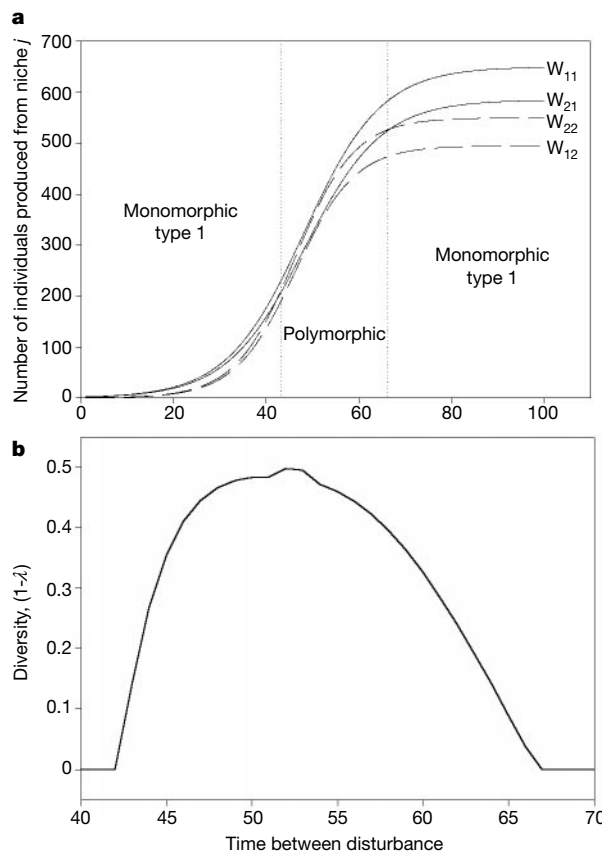


Figure 2 The effect of disturbance frequency on diversity, in the case where competitive ability and disturbance tolerance are positively related. Parameter values are: type 1 in niche 1 (W_{11}), $r = 0.12$, $K = 650$, $T = 0$; type 2 in niche 2 (W_{22}), $r = 0.155$, $K = 550$, $T = 10$. **a**, The number of individuals of type j produced from niche j as a function of time between disturbance. **b**, The behaviour of $1 - \lambda$ at equilibrium.

been allowed to diversify, whereas in the second the founding populations were genetically uniform, comprising a single ancestral SM genotype. Disturbances were nonspecific mass mortality events: after thorough homogenization, 99.9% of the culture was discarded and the remaining 0.1% ($\sim 10^7$ bacteria) transferred to fresh media-filled microcosms. The populations thus went through a minimum of 10 generations ($\log_2(1,000)$) between disturbances. Experiments were performed over a 16-day period in both shaken (homogeneous) and static (heterogeneous) microcosms under five different disturbance regimes: daily, every second day, fourth day, eighth day, or not at all during the course of 16 days. After 16 days, cultures were plated onto agar and diversity determined by scoring the frequency of colony morphologies.

Under all disturbance conditions, diversity was two to three times greater in the static than in the shaken environment (Fig. 3), supporting the main result of previous work—that spatial heterogeneity is crucial both to the emergence and to the maintenance of diversity¹⁵. Intermediate frequencies of disturbance caused a further 30% increase in diversity in the heterogeneous microcosms, resulting in a unimodal relationship between diversity and transfer frequency (Fig. 3a, linear term, $F_{1,78} = 0.84$, $P > 0.1$; negative quadratic term, $F_{1,77} = 44.68$, $P < 0.001$). This was not the case in the homogenous microcosms, where this relationship was negative and monotonic (Fig. 3b, linear term, $F_{1,77} = 29.7$, $P < 0.001$; quadratic term, $F_{1,76} = 0.19$, $P > 0.1$). In both heterogeneous and homogeneous microcosms the relationship between diversity and disturbance frequency was the same, whether the founding cultures were initially uniform or diverse ($P > 0.1$ for the interaction between disturbance frequency and base population).

Heterogeneous microcosms all contained variants of both the broth-colonizing SM and air–broth interface-colonizing WS

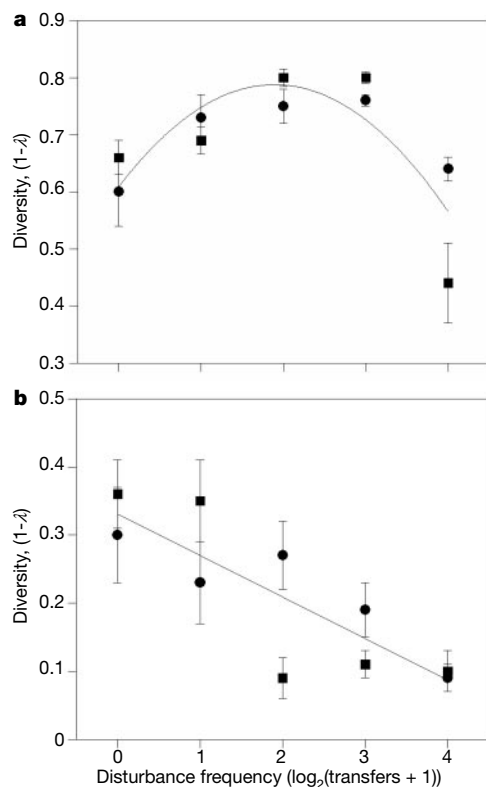


Figure 3 Mean diversity ($1 - \lambda$) at different frequencies of disturbance in heterogeneous and homogeneous environments, after 16 days. Values of the mean are given as ± 1 s.e.m., $n = 8$. **a**, Heterogeneous environment; **b**, homogeneous environment. Cultures were initiated with either isogenic (circles) or diverse (squares) base populations.

morphs; some contained a small proportion ($< 1\%$) of FS morphs. Cultures disturbed at intermediate frequencies contained approximately equal numbers of SM and WS morphs, whereas at high or low disturbance SM or WS morphs dominated in different replicates. Only variants of the ancestral-like SM morph were present in the homogeneous microcosms. The negative relationship between diversity and disturbance frequency in these microcosms (Fig. 3b) is probably an effect of the number of generations and hence the opportunity for periodic selection to purge transient variation. Asexual populations in homogeneous environments are characterized by periodic selective sweeps^{21,22}, therefore clonal replacement would occur more slowly in undisturbed populations (fewer generations) revealing more diversity.

Differently adapted specialists can coexist if each has greater fitness than the other when rare (negative frequency-dependent selection)^{23,24}. If stable coexistence is more likely at intermediate disturbance frequencies, then competing WS and SM genotypes should both be more likely to invade a population dominated by the other type under this regime. To test this, 12 independent pairs of SM and WS genotypes, with opposing genetic markers, were cultured as mixtures. The frequency of the rare genotype was 0.01 in each case, and populations were cultured for 16 days under high, medium and low disturbance frequencies (daily, 4-daily and no disturbance over a 16-day period, respectively). The frequencies of genotypes were scored on agar plates and used to calculate relative fitness. Before conducting this experiment we confirmed that diversity in the paired SM–WS populations was indeed greatest at

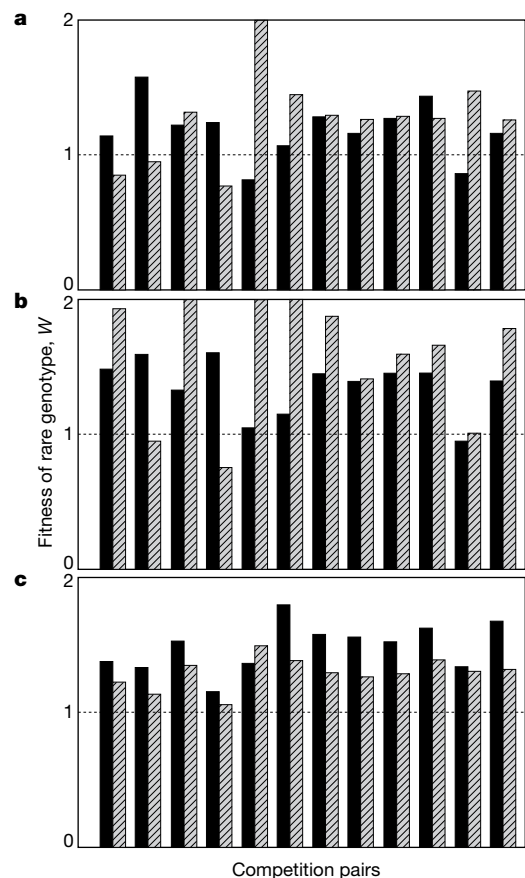


Figure 4 Relative fitness (W) of WS (solid bars) and SM (hatched bars) genotypes after 16 days when initiated at a ratio of 1:100. Each pair of bars shows the fitness when rare of the WS and SM genotypes in each of the 12 competition pairs. The same genotype pairs were competed under daily transfer (**a**), no transfer (**b**) and 4-daily transfer (**c**) conditions. The horizontal line across each graph indicates equal fitness of the rare and common genotypes in each competition ($W = 1$).

intermediate disturbance ($F_{2,33} = 5.20$, $P < 0.01$), and that diversity at high and low disturbance were comparable (effect of combining high and low disturbance treatments: $F_{1,33} = 1.18$, $P > 0.1$).

Under the intermediate disturbance regime the rare genotype increased in frequency relative to the common genotype in every instance (Fig. 4). However, in populations subjected to either high or low disturbance only certain genotypes were able to invade when rare. We failed to observe the operation of negative frequency-dependent selection in five populations subjected to high disturbance and in three populations subjected to low disturbance (Fig. 4). Moreover, the mean variance in fitness of each WS and SM pair was significantly lower under intermediate (0.03), than high (0.11) or low (0.16), disturbance frequencies (intermediate versus high: $F_{2,2} = 3.67$, $P < 0.05$; Fig. 4).

The original formulation of the intermediate disturbance hypothesis requires a trade-off between competitive ability and disturbance tolerance⁸. Although this is not a requirement of our model, such a trade-off might have affected diversity in our experimental microcosms. We investigated this possibility by calculating genetic correlations for fitness of the WS and SM morphs under the high and low disturbance regimes. In 10 of the 12 populations, the same genotype in each pair was fittest under both the high and low disturbance regimes (sign test of correlation coefficients: $P < 0.05$; Fig. 4a and b) showing that no general trade-off between disturbance tolerance and competitive ability was expressed.

Explanations for the unimodal relationship between disturbance and diversity are typically cast in terms of local patch dynamics, and there is good evidence from field studies^{4,5} that this is often an appropriate interpretation. Many natural ecosystems are subject to disturbances that act more or less uniformly over large areas, such as grazing or fires. It has long been known that grazing may change the species composition of the community²⁵, and in many cases grazed areas have higher species diversity²⁶. Moreover, experimental removal of tissue by mowing causes an increase in diversity similar to that associated with grazing by either exotic²⁷ or native²⁸ herbivores. The mechanism we have identified here may explain why such uniform disturbances can lead to increased diversity. Disturbance, like productivity, can modulate the effect of environmental heterogeneity by changing the relative production¹⁶ of different niches. □

Methods

Disturbance in heterogeneous and homogeneous environments

Cultures were initiated with 10^7 cells from either an isogenic or diverse population of *P. fluorescens* strain SBW25 (ref. 15), and cultured at 28°C in 25-ml glass universals with loose plastic lids, containing 6 ml of King's Medium B (KB). An isogenic population was produced by growing ancestral SBW25 for 18 h under shaken conditions. Two diverse populations were created by culturing ancestral SBW25 for 4 d under shaken and static conditions. The shaken and static diverse populations were used to initiate experimental cultures subsequently exposed to shaken and static conditions, respectively. Five disturbance frequencies were used: 6 µl of culture was transferred to fresh medium every day, every second, fourth and eighth day, and not at all, during a 16-day period. Four isogenic and 4 diverse replicate populations were established for each treatment in each environment (shaken and static). The experiment was replicated, but with different starting populations. Diversity was measured as the complement of Simpson's index of concentration ($1 - \lambda$), calculated from approximately 100 colonies grown on KB agar plates: $\lambda = \sum p_i^2$, where p_i is the proportion of the i th morph²⁹.

Competition experiments

Six SM and 6 WS genotypes were isolated from 4-day-old cultures grown in static tubes established from both the wild type and an SBW25 pantothenate-requiring auxotroph¹⁵. Each SM genotype was competed against an independent WS genotype with a different genetic background at 100:1 and 1:100 ratios (a total of 10^5 cells per ml starting populations) under the three disturbance regimes in static tubes containing 6 ml KB medium supplemented with 0.0024% pantothenic acid, negating any fitness cost of the panB mutation¹⁵. Diversity ($1 - \lambda$) was calculated from the frequency of wild-type and panB genotypes, and averaged across the two reciprocal competition cultures. Relative fitness (W ; Fig. 4) was calculated from the ratio of the estimated malthusian parameters

(m) of the rare: common genotypes²⁰, $m = \ln(N_f/N_0)$, where N_0 is the starting density and N_f the final density (determined by counting at least 2,000 and 200 colonies, respectively grown on vitamin-free KB agar supplemented with $4.8 \times 10^{-6}\%$ pantothenic acid—on this medium the pantothenate marked strain is readily distinguished by its greatly reduced size).

Statistical analyses

Analyses were performed by stepwise deletion from full statistical models using GLIM 4; non-significant terms ($P > 0.1$) were included in the error variance when calculating test statistics of remaining terms³⁰. The effect of disturbance frequency on (logit-transformed) diversity in Fig. 3 was analysed separately for cultures in the shaken and non-shaken environments. Experiment number (1 or 2) and base population (isogenic or diverse) were fitted as 2-level factors, and $\log_2$ (transfer rate + 1) fitted as both a linear and quadratic covariate. For (logit-transformed) diversity in the competition cultures, both disturbance frequency (high, low and intermediate) and genotype pair were fitted as factors. Orthogonal contrasts were performed to determine whether diversity under high and low disturbance differed³⁰.

Received 14 August; accepted 23 October 2000.

- Grime, J. P. Control of species diversity in herbaceous vegetation. *J. Environ. Mgmt* **1**, 151–167 (1973).
- Connell, J. H. Diversity in tropical rain forests and coral reefs. *Science* **199**, 1302–1310 (1978).
- Huston, M. A. *Biological Diversity* (Cambridge Univ. Press, Cambridge, 1994).
- Floder, S. & Sommer, U. Diversity in planktonic communities: An experimental test of the intermediate disturbance hypothesis. *Limnol. Oceanogr.* **44**, 1114–1119 (1999).
- Sousa, W. P. Disturbance in marine intertidal boulder fields: the nonequilibrium maintenance of species diversity. *Ecology* **60**, 1225–1239 (1979).
- Gaedeke, A. & Sommer, U. The influence of the frequency of periodic disturbances on the maintenance of phytoplankton diversity. *Oecologia* **71**, 25–28 (1986).
- Weider, L. J. Disturbance, competition and the maintenance of clonal diversity in *Daphnia pulex*. *J. Evol. Biol.* **5**, 505–522 (1992).
- Petratis, P. S., Latham, R. E. & Niesenbaum, R. A. The maintenance of species diversity by disturbance. *Q. Rev. Biol.* **64**, 393–418 (1989).
- Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995).
- Slatkin, M. Competition and regional coexistence. *Ecology* **55**, 128–134 (1974).
- Levin, S. A. & Paine, R. T. Disturbance, patch formation, and community structure. *Proc. Natl Acad. Sci. USA* **71**, 2744–2747 (1974).
- Chesson, P. & Huntly, N. The roles of harsh and fluctuating conditions in the dynamics of ecological communities. *Am. Nat.* **150**, 519–553 (1997).
- Korona, R., Nakatsu, C. H., Forney, L. J. & Lenski, R. E. Evidence for multiple adaptive peaks from populations of bacteria evolving in a structured habitat. *Proc. Natl Acad. Sci. USA* **91**, 9037–9041 (1994).
- Bell, G. A. C. Experimental evolution in *Chlamydomonas*. 1. Short-term selection in uniform and diverse environments. *Heredity* **78**, 490–497 (1997).
- Rainey, P. B. & Travisano, M. Adaptive radiation in a heterogeneous environment. *Nature* **394**, 69–72 (1998).
- Kassen, R., Buckling, A., Bell, G. & Rainey, P. B. Diversity peaks at intermediate productivity in experimental microcosms. *Nature* **406**, 508–512 (2000).
- Levene, H. Genetic polymorphism when more than one ecological niche is available. *Am. Nat.* **87**, 331–333 (1953).
- Gause, G. F. *The Struggle for Existence* (Williams and Wilkins, Baltimore, 1934).
- Rainey, P. B., Buckling, A., Kassen, R. & Travisano, M. The emergence and maintenance of diversity: Insights from experimental bacterial populations. *Trends Ecol. Evol.* **15**, 243–247 (2000).
- Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-term experimental evolution in *Escherichia coli*. 1. Adaptation and divergence during 2,000 generations. *Am. Nat.* **138**, 1315–1341 (1991).
- Atwood, K. C., Schneider, L. K. & Ryan, F. J. Selective mechanisms in bacteria. *Cold Spring Harb. Symp. Quant. Biol.* **16**, 345–354 (1952).
- Elena, S. F., Cooper, V. S. & Lenski, R. E. Punctuated evolution caused by selection of rare beneficial mutations. *Science* **272**, 1802–1804 (1996).
- Haldane, J. B. S. *The Causes of Evolution* (Harper, New York, 1932).
- Ayala, F. J. & Campbell, C. A. Frequency-dependent selection. *Annu. Rev. Ecol. Syst.* **5**, 115–138 (1974).
- Milton, W. E. J. & Davies, R. O. The yield, botanical and chemical composition of natural hill herbage under manuring, controlled grazing and hay conditions. *J. Ecol.* **35**, 65–95 (1947).
- Milchunas, D. G. & Lauenroth, W. K. Quantitative effects of grazing on vegetation and soils over a global range of environments. *Ecol. Monogr.* **63**, 327–366 (1993).
- Collins, S. L. Interaction of disturbances in tallgrass prairie: a field experiment. *Ecology* **68**, 1243–1250 (1987).
- Collins, S. L., Knapp, A. K., Briggs, J. M., Blair, J. M. & Steinauer, E. M. Modulation of diversity by grazing and mowing in native tallgrass prairie. *Science* **280**, 745–747 (1998).
- Simpson, E. H. Measurement of diversity. *Nature* **163**, 688 (1949).
- Crawley, M. J. *GLIM for Ecologists* (Blackwell Science, Oxford, 1993).

Acknowledgements

We thank J. Pannell, N. Colegrave, K. McCann and M. Brockhurst for comments and discussion. This work was supported by the UK NERC and BBSRC (A.B. and P.B.R.), the Canadian NSERC (G.B. and R.K.) and the British Council.

Correspondence and requests for materials should be addressed to A.B. (e-mail: angus.buckling@plant-sciences.ox.ac.uk).

Altruism and social cheating in the social amoeba *Dictyostelium discoideum*

Joan E. Strassmann, Yong Zhu & David C. Queller

Department of Ecology and Evolutionary Biology, Rice University, PO Box 1892, Houston, Texas 77251-1892, USA

The social amoeba, *Dictyostelium discoideum*, is widely used as a simple model organism for multicellular development^{1,2}, but its multicellular fruiting stage is really a society. Most of the time, *D. discoideum* lives as haploid, free-living, amoeboid cells that divide asexually. When starved, 10⁴–10⁵ of these cells aggregate into a slug. The anterior 20% of the slug altruistically differentiates into a non-viable stalk, supporting the remaining cells, most of which become viable spores^{3–5}. If aggregating cells come from multiple clones, there should be selection for clones to exploit other clones by contributing less than their proportional share to the sterile stalk. Here we use microsatellite markers to show that different clones collected from a field population readily mix to form chimaeras. Half of the chimaeric mixtures show a clear cheater and victim. Thus, unlike the clonal and highly cooperative development of most multicellular organisms, the development of *D. discoideum* is partly competitive, with conflicts of interests among cells. These conflicts complicate the use of *D. discoideum* as a model for some aspects of development, but they make it highly attractive as a model system for social evolution.

The problem of potential conflict among multiple clones within fruiting bodies has been recognized^{6–13}, but little investigated. The altruistic stalk cells give up reproduction in order to benefit the spore cells, by lifting them above the hazards of the soil or increasing their chances of dispersal to a more favourable environment^{7,14}. When fruiting bodies contain only one clone, like multicellular animals, sterile stalk cells are favoured by kin selection¹⁵; genes for facultative sterility can spread if they help copies of themselves in the reproductive cells. However, if *D. discoideum* aggregates include more than one clone, they may be more analogous to the societies of social insects, which have sterile castes, but also have multiple genotypes and genetic conflicts of interests⁷. Selection within individual fruiting bodies could favour cheater clones that contribute less than their fair share of the sterile stalk^{6–13}. Such conflicts would be averted if Dictyostelid clones always multiply and aggregate in isolation from other clones^{6,8} but, in species where it has been examined, genetically diverse clones do sometimes occur in close proximity in nature^{16,17}.

Within-individual selection and developmental conflict could be minimized in two other ways. First, cells might recognize and exclude foreign clones⁹, as occurs in many marine invertebrates¹⁸. If this does not occur, there might still be some mechanism that assures fairness in chimaeric mixtures⁹ analogous to the way meiosis normally ensures fairness in segregation. There has been no systematic study of these points in Dictyostelids, particularly in natural populations. Chimaeras are sometimes generated, usually with laboratory strains^{19–21}. However, recognition abilities could have been lost in a laboratory environment where they are not needed. Moreover, because many of these laboratory mixtures involve a developmental mutant and its otherwise isogenic parent clone, they would not reveal a genetic recognition system based on genetic differences at other loci. Several such chimaeras show unequal apportionment of the two clones to spore and stalk^{19–21}, but it is not known if such cheating is common in nature. Buss⁹ found a natural stalkless clone of *Dictyostelium mucoroides* that could exploit the stalk-forming ability of a neighbouring clone. It did

not mix with other clones in the vicinity, suggesting that the parasite might have been a new mutant that could only victimize its parent.

To test whether clones generally recognize and exclude each other, we obtained clonal isolates of *D. discoideum* collected from a natural population at the type locality in North Carolina. We performed pairwise mixing experiments, testing for chimaeric slugs by amplifying microsatellite loci whose alleles differed in the two clones. The presence of both parental alleles in a slug indicates chimaera formation (Fig. 1). All 15 pairs of clones tested consistently formed chimaeras. This failure to exclude non-relatives renders *D. discoideum* clones potentially vulnerable to cheating.

A second experiment shows that chimaeric mixtures are often not fair, in the sense that the clonal composition of spores differs from that of stalks. We made 12 mixtures of two clones each. Instead of genotyping entire slugs, we sampled the prespore and prestalk regions of each slug, using the fact that the anterior 20% of migrating slugs will form the stalk. For each sample, we amplified a variable microsatellite locus, ran it out on a gel to separate the two distinct parental alleles, and quantified the amount of the two alleles. Because the two alleles are amplified in the same reaction and have the same primer binding sites, the polymerase chain reaction (PCR) products should be proportional to the relative numbers of target sequences, that is, to the relative cell numbers of the two clones. These relative counts are sufficient to test the null hypothesis of fair behaviour, which is that a clone's representation in the stalk should be the same as its representation in the spores. For example, if a clone contributes 70% of the spores, it should also contribute 70% of the stalk. We note that we do need both numbers; 70% representation in spores is not by itself evidence for cheating because it could be due to other factors like differential growth rates before fruiting. The null hypothesis is rejected in 6 of the 12 mixtures (Fig. 2), indicating that cheating by one clone is common.

The results so far are consistent with two forms of cheating. They appear to suggest that one clone often gains by producing more spore cells than it normally would, exploiting the other clone's stalk cells. However, unequal representation in the stalk and spores could also arise from different solitary allocations, unchanged in mixture. For example, if clone A always has a stalk/spore allocation of 30/70 and clone B always has 10/90, in a 50:50 mixture clone A will contribute only 1/4 (that is, 10/(10+30)) of the stalk cells, while obtaining 9/16 (that is, 90/(70+90)) of the spores. In this fixed allocation model, the cheater gets no more spores in mixture than it does alone, but it may gain another advantage. It will have its spores at a greater height in mixture than alone. Similarly, its victim clone will have its spores on a shorter stalk in mixture than alone. The cheater's cost-free increase in stalk height is likely to be advantageous, provided stalk heights have been previously selected to reflect

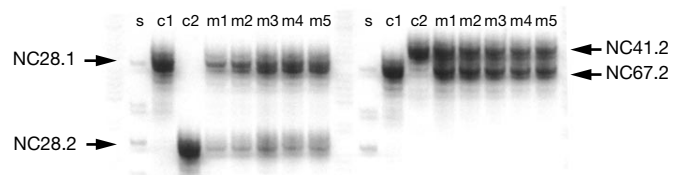


Figure 1 Microsatellite genotypes of slugs from two-clone mixture experiments. For each experiment, lanes c1 and c2 are slugs from the two clones cultured alone (labelled at the side with the clone name), each showing a single allele, as expected for haploid *D. discoideum*. Lanes m1–m5 show five slugs derived from mixing amoebae of the two clones. Each is chimaeric, showing both alleles. Lane s is a size marker. All 15 clone pairs tested showed clear mixing, with an average of 21 slugs tested. (When contributions of the two parent clones appeared unequal, as for NC 28.1 and NC 28.2 on the left, the inequalities tended to be similar across slugs, consistent with differences in growth rates before aggregation, or with differences in size or viability of the sori used to initiate the mixture.)

a balance between benefit and cost. When stalks are costly, their average length would be set below the height that would yield maximum benefits to spores. Cheaters gain such height benefits by allowing the victims to pay more of the costs. Buss's⁹ stalkless mutant was apparently a cheater of this form, allocating no cells to stalk when alone or in mixture.

Distinguishing between these two forms of cheating requires estimates of solitary allocation to spores and stalks. Direct cell counts are not feasible for stalks, so we used two different indices of relative allocation. For the first, we measured 15 randomly chosen fruiting bodies of each clone. As a measure of a clone's relative allocation to spores, we took the mean of S_1 , where $S_1 = (\text{sorus width})/(\text{sorus width} + \text{stalk length})$, where the sorus is the spherical cluster of spore cells on top of the stalk. Our second relative estimator of spore allocation, S_2 , was the length of a slug's prespore region divided by the length of the slug. Each clone's mean was estimated from 20 randomly chosen slugs of each clone stained with neutral red, which dyes the anterior prestalk region. For each of the mixtures in Fig. 2, we then calculated a measure of how much

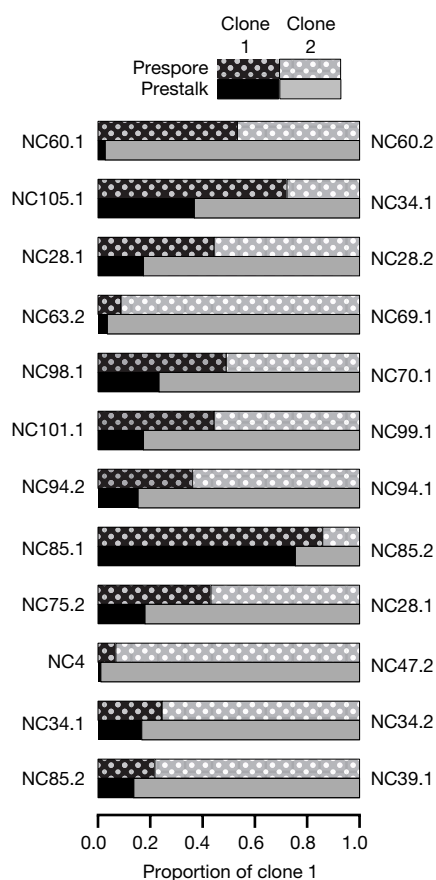


Figure 2 Relative contributions of mixed clones to prespore and prestalk regions of chimaeras. Each of the 12 four-pattern bars shows the averages from seven replicates of one pairwise mixture of two clones. The legend at the top shows that the top half (speckled) represents the prespore region and the bottom half (solid) represents the prestalk region. Each of these is divided according to the contributions of the clone on the left (dark) and the clone on the right (light). For example, in the mixture of clones 60.1 and 60.2, clone 60.1 gets 53% of the prespore cells (dark speckled), but contributes only 3% of the prestalk (dark solid). This disparity makes 60.1 a cheater, gaining spores without paying its share of the cost. The null hypothesis—that a clone contributes equal percentages to the prespore and prestalk regions—would be represented by equal-sized dark portions in both top and bottom bars, as in the legend at the top. It is rejected for the top nine mixtures taken individually (t -tests, paired by slug, of arcsine-squareroot transformed proportions), and for the top six mixtures when corrected for multiple comparisons (step-down Holm adjustment)³⁰.

more (or less) the selfish clone allocates to spores than its victim does, when each is alone: $D_1 = S_1(\text{selfish}) - S_1(\text{victim})$, and $D_2 = S_2(\text{selfish}) - S_2(\text{victim})$. The fixed allocation hypothesis predicts that these measures should be positive (the more “spory” clone should be the selfish one in mixture), but D_1 was positive in only 4 of 12 cases, and D_2 in 6 of 12. It also predicts that D_1 and D_2 should be positively related to the magnitude of selfishness in mixture (as measured by the selfish clone's spore proportion minus its stalk proportion in Fig. 2). The prediction fails; both regressions are strongly and significantly negative (both slopes = 0.59, $P < 0.05$) indicating that typically higher allocation to spores alone means less selfishness in mixture. Thus, there is no evidence that cheating is restricted to the fixed-allocation advantage of having a higher stalk. Instead, it must often involve spore number advantages. Certainly fixed allocation cannot explain extreme results such as the 60.1/60.2 mixture (Fig. 2); clone 60.1's allocation in mixture would imply that it has virtually no stalk when alone, but in fact it has a normal stalk.

The ability to cheat appears to be common in this natural population, something that has not been demonstrated for other examples of developmental cheating in a social bacterium²², a colonial ascidian²³, and *Dictyostelium*^{9,20,21}. Most previous examples come from laboratory mutants^{20,21,22} whose significance in natural populations is unknown. In examples from natural populations, the incidence of cheating appears low: ascidians usually exclude each other²⁴, and only one natural stalkless mutant of *D. mucoroides* has been found⁹. How cheating is maintained at such high levels in *D. discoideum* is not yet known. Factors that may be important are the frequency of between-clone encounters in the field, whether a clone's ability to cheat is general or specific to certain victim clones, and costs of being a cheater.

The prevalence of cheating in *D. discoideum* complicates its use as a model system for development of organisms like vertebrates, which are clonal and essentially free from such conflicts. Within-cell processes may be little affected by this difference, but between-cell interactions involving signalling, adhesion and differentiation may be very different. Signalling among non-relatives often involves elements of deception and manipulation²⁵, as when female spiders performing courtship behaviours eat the male instead of mating with him²⁶. Recognizing that between-cell signalling in *D. discoideum* may have similar conflict elements might help resolve the long-standing problem of which cells go to spore and stalk, much as recognition of worker–queen conflict was essential for understanding social insect sex ratios²⁷. These conflicts of interest might even account for the very feature that has made *Dictyostelium* attractive as a model developmental system: its simplicity. Social amoebae appear to possess most of the kinds of molecular mechanisms required to evolve more complex forms of multicellularity, but they have not done so²⁸. Perhaps what they lack most is an effective means to suppress conflict, either by assuring genetic uniformity or by enforcing fairness in mixtures.

Biologists interested in social evolution have long been hampered by the difficulties of studying organisms like social insects. Selection experiments are rarely feasible, and there is little foundation for identifying the genes that affect sociality and tracing their effects. In contrast, such studies should be facilitated in *D. discoideum* by its short generation time, the imminent completion of the genome sequence²⁹, and decades of careful work on its cell biology and its molecular biology, making it an invaluable model system for evolutionary and mechanistic studies of altruism and cheating. □

Methods

We obtained natural clonal isolates of *D. discoideum* collected from the type locality, Little Butts Gap, North Carolina. These were collected from a 100 m × 50 m wooded area (clones 26–75) or within 1,500 m of this area (clones 76–105) and had previously been shown to be genetically diverse, sometimes within the same small soil sample¹⁶. We also used the original type clone, NC4.

Microsatellite sequences, which are tandem repeats of short DNA motifs, were located

in GenBank and from the Genome Sequencing Centre Jena (<http://genome.imb-jena.de/dictyostelium/>). Primers were designed to amplify the repeat sequence (information is available from the authors on request), which is often variable in length.

To determine if clones mix, we performed 15 pairwise mixing experiments. For each, we grew up fruiting bodies for the two clones separately, picked 5 sori from each, and mixed all 10 with *Klebsiella aerogenes* as a food source in 0.2 ml of sterile distilled water, where they dispersed as hundreds of thousands of independent amoebae. We spread this homogeneous mixture of two clones out on an SM/5 plate. After several days growth at 22 °C, the starving amoebae aggregated and formed migrating slugs, and 24 were chosen for genotyping at a microsatellite locus chosen because the two parent clones possessed different alleles. Mixing was confirmed if a slug showed both parental alleles. We mixed all the clone pairs listed in Fig. 2 except NC94.1-NC94.2 (which did not grow in this trial), plus four others: NC41.2-NC67.2; NC54.1-NC54.2; NC42.1-NC43.1; and NC66.2-NC85.1. We use the original nomenclature¹⁶.

To assess the frequency of cheating in mixtures, we made 12 pairwise mixtures (listed in Fig. 2) as in the first experiment. For each mixture, we selected 7 migrating slugs. The front 10% of the slug was excised as a sample of the prestalk region. The prespore region of each slug was sampled twice, using the posterior 10% of the slug and an equivalent portion from the middle of the prespore region. We extracted DNA separately from all three portions and amplified a microsatellite locus that differed for the two parent clones, incorporating ³⁵S-labelled dATP in the PCR reaction. We ran the product DNA out on a 6% polyacrylamide gel. We then quantified the relative amounts of the two parental alleles using a phosphorimager that measures the radiation given off by any region of the gel. To remove background, we subtracted the average radiation reading of two other areas, located above and below the bands, and identical in size to the band. Exposures were kept low to avoid saturation effects. We compared the prestalk with the average of the two prespore samples. A replicate set of the same 12 mixtures was conducted, using only the posterior 10% of the slug for the prespore region, and 11 of the 12 showed differences in the same directions as in Fig. 2.

Marker loci were chosen such that the two alleles did not have overlapping stutter bands, but were within a few repeats of each other to avoid differential amplification. However, even if one of the two alleles amplified better, it would do so consistently in prestalk and prespore samples, so that a test for a difference between these two samples remains valid. We confirmed that the method works by amplifying a second locus for five of our mixtures. For each, the correlation across slugs of two locus-specific estimates of %prestalk minus %prespore was highly significant ($P < 0.005$) with the correlation averaging 0.8. If the results we reported were due to differential amplification or some other artefact, we would expect no correlation, or negative correlations as often as positive.

Received 15 September; accepted 18 October 2000.

- Maeda, Y., Inouye, K. & Takeuchi, I. (eds) *Dictyostelium—A Model System for Cell and Developmental Biology* (Universal Academy, Tokyo, 1997).
- Gross, J. D. Developmental decisions in *Dictyostelium discoideum*. *Microbiol. Rev.* **58**, 330–351 (1994).
- Bonner, J. T. *The Cellular Slime Molds* (Princeton Univ. Press, Princeton, 1967).
- Raper, K. B. *The Dictyostelids* (Princeton Univ. Press, Princeton, 1984).
- Raper, K. B. Pseudoplasmodium formation and organization in *Dictyostelium discoideum*. *Journal of the Elisha Mitchell Scientific Society* **56**, 241–282 (1940).
- Williams, G. C. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* (Princeton Univ. Press, Princeton, 1966).
- Gadagkar, R. & Bonner, J. T. Social insects and social amoebae. *J. Biosci.* **19**, 219–245 (1994).
- Armstrong, D. P. Why don't cellular slime molds cheat. *J. Theor. Biol.* **109**, 271–283 (1984).
- Buss, L. W. Somatic cell parasitism and the evolution of somatic tissue compatibility. *Proc. Natl Acad. Sci. USA* **79**, 5337–5341 (1982).
- Hilson, J. A., Kolmes, S. A. & Nellis, L. F. Fruiting body architecture, spore capsule contents, selfishness, and heterocytosis in the cellular slime mold, *Dictyostelium discoideum*. *Ethol. Ecol. Evol.* **6**, 529–535 (1994).
- Wilson, D. S. & Sober, E. Reviving the superorganism. *J. Theor. Biol.* **136**, 337–356 (1989).
- Kessin, R. H. in *Dictyostelium—A Model System for Cell and Developmental Biology* (eds Maeda, Y., Inouye, K. & Takeuchi, I.) 3–13 (Universal Academy, Tokyo, 1997).
- Atzmony, D., Zahavi, A. & Nanjundiah, V. Altruistic behaviour in *Dictyostelium discoideum* explained on the basis of individual selection. *Curr. Sci.* **72**, 142–145 (1997).
- Huss, M. J. Dispersal of cellular slime moulds by two soil invertebrates. *Mycologia* **81**, 677–682 (1989).
- Hamilton, W. D. The genetical evolution of social behaviour. I. *J. Theor. Biol.* **7**, 1–52 (1964).
- Francis, D. & Eisenberg, R. Genetic structure of a natural population of *Dictyostelium discoideum*, a cellular slime mold. *Mol. Ecol.* **2**, 385–391 (1993).
- Ketcham, R. B. & Eisenberg, R. M. Clonal diversity in populations of *Polysphondylium pallidum*, a cellular slime mold. *Ecology* **70**, 1425–1433 (1989).
- Grosberg, R. K. The evolution of allorecognition specificity in clonal invertebrates. *Q. Rev. Biol.* **63**, 377–412 (1988).
- Filosa, M. F. Heterocytosis in cellular slime molds. *Am. Nat.* **96**, 79–91 (1962).
- Houle, J., Balthazar, J. & West, C. M. A glycosylation mutation affects cell fate in chimeras of *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci. USA* **86**, 3679–3683 (1989).
- Ennis, H. L., Dao, D. N., Pukatzki, S. U. & Kessin, R. H. *Dictyostelium* amoebae lacking an F-box protein form spores rather than stalk in chimeras with wild type. *Proc. Natl. Acad. Sci. USA* **97**, 3292–3297 (2000).
- Velicer, G. J., Kroos, L. & Lenski, R. E. Developmental cheating in the social bacterium *Myxococcus xanthus*. *Nature* **404**, 598–601 (2000).
- Stoner, D. S., Rinkevich, B. & Weissman, I. L. Heritable germ and somatic cell lineage competitions in chimeric colonial protochordates. *Proc. Natl. Acad. Sci. USA* **96**, 9148–9153 (1999).
- Grosberg, R. K. & Quinn, J. F. The genetic control and consequences of kin recognition by the larvae of a colonial marine invertebrate. *Nature* **322**, 456–459 (1986).
- Dawkins, R. & Krebs, J. R. in *Behavioural Ecology: An Evolutionary Approach* 2nd edn (eds Krebs, J. R. & Davies, N. B.) 282–309 (Blackwell, Oxford, 1978).

- Elgar, M. A. & Nash, D. R. Sexual cannibalism in the garden spider *Araneus diadematus*. *Anim. Behav.* **36**, 1511–1517 (1988).
- Queller, D. C. & Strassmann, J. E. Kin selection and social insects. *Bioscience* **48**, 165–175 (1998).
- Loomis, W. F. *Four Billion Years; An Essay on the Evolution of Genes and Organisms* (Sinauer, Sunderland, Massachusetts, 1988).
- Kay, R. R. & Williams, J. G. The *Dictyostelium* genome project—an invitation to species hopping. *Trends Genet.* **15**, 294–297 (1999).
- Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* **6**, 65–70 (1979).

Acknowledgements

We thank R. Gomer for advice and training; D. Welker for supplying the clones; J. Keay, W. Castle, S. Reddy and J. Damon for assistance with laboratory work; and D. Rozen, R. Kessin, G. Velicer and J. Bonner for comments on the manuscript. This work was supported in part by the US National Science Foundation.

Correspondence and requests for materials should be addressed to J.E.S. (e-mail: strassm@rice.edu).

Asymmetric leaves1 mediates leaf patterning and stem cell function in Arabidopsis

Mary E. Byrne*, Ross Barley†, Mark Curtis*, Juana Maria Arroyo*, Maitreya Dunham*, Andrew Hudson† & Robert A. Martienssen*

* Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA

† Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, UK

Meristem function in plants requires both the maintenance of stem cells and the specification of founder cells from which lateral organs arise. Lateral organs are patterned along proximodistal, dorsoventral and mediolateral axes^{1,2}. Here we show that the *Arabidopsis* mutant *asymmetric leaves1 (asl1)* disrupts this process. *ASL1* encodes a myb domain protein, closely related to *PHANTASTICA* in *Antirrhinum* and *ROUGH SHEATH2* in maize, both of which negatively regulate knotted-class homeobox genes. *ASL1* negatively regulates the homeobox genes *KNAT1* and *KNAT2* and is, in turn, negatively regulated by the meristematic homeobox gene *SHOOT MERISTEMLESS*. This genetic pathway defines a mechanism for differentiating between stem cells and organ founder cells within the shoot apical meristem and demonstrates that genes expressed in organ primordia interact with meristematic genes to regulate shoot morphogenesis.

The shoot apical meristem (SAM) comprises slowly dividing stem cells at the centre and daughter cells at the periphery, from which organ founder cells are recruited. Founder cells divide rapidly, initiating the outgrowth of organ primordia while polarity is established along the proximodistal, dorsoventral and mediolateral axes^{1,2}. The mechanism by which stem cell and founder cell derivatives are distinguished is obscure, but is likely to involve a highly conserved class of homeobox genes related to *KNOTTED1* in maize (*KNOX* genes). *KNOX* genes are expressed in the SAM but are downregulated in founder cells at the time of leaf initiation³. They are implicated in maintaining division or preventing differentiation of cells in the SAM. Loss-of-function mutations in the *Arabidopsis* *KNOX* gene *SHOOT MERISTEMLESS (STM)* result in embryos that lack a SAM^{4,5}. Recessive mutations in the *kn1* gene of maize are also defective in meristem maintenance⁶. In contrast, gain-of-function mutations that result in ectopic expression of *KNOX* genes in maize disrupt normal leaf development causing distal displacement

* Present address: Institute of Biological Sciences, University of Wales, Aberystwyth, UK (M.C.); Genetics Department, Stanford, California 94305, USA (M.D.).

of sheath and auricle tissue into the blade⁷. In dicotyledonous plants, such as *Arabidopsis* and tobacco, misexpression of homeobox transgenes results in lobed leaves, distally displaced stipules, and occasional ectopic meristems on the adaxial leaf surface^{8,9}.

PHANTASTICA (*PHAN*) in *Antirrhinum* is transcribed in organ founder cells and is required to prevent expression of an *STM*-like *KNOX* gene in lateral organs^{10,11}. *phan* mutants show variable defects in leaf patterning, affecting both dorsoventral and proximodistal axes¹². *PHAN* has a myb domain found in several transcription factors and its homologue in maize is *ROUGH SHEATH2* (*RS2*)^{11,13}. Mutations in *RS2* result in phenotypes comparable to dominant mutations in *KNOX* genes, and show ectopic expression of homeodomain proteins in developing leaves revealing a conserved role for *RS2* in repressing *KNOX* gene expression^{11,13,14}.

asymmetric leaves1 (*as1*) is a classical mutation in *Arabidopsis* that disrupts development of cotyledons, leaves and floral organs (Fig. 1)¹⁵. The mutant leaf lamina is subdivided into prominent outgrowths or lobes. Early leaves have few lobes located proximally, whereas later leaves have more lobes extending distally toward the leaf tip (Fig. 1c). Adult rosette leaves of *magnifica*, the first isolated mutant allele of *as1* (ref. 16), have patches of callus-like growth on the leaf lamina, and occasionally have ectopic shoots on the adaxial surface of the petiole (Fig. 1d), similar to those observed in

transgenic plants misexpressing the *KNOX* gene *KNAT1* (ref. 9). In wild-type leaves the abaxial epidermis over the midvein of the leaf petiole and blade is comprised of elongated cells (Fig. 1e). In *as1* there are multiple bundles of elongated cells extending from the petiole into the blade (Fig. 1f). The enhancer trap (see Methods) ET2689 marks these elongated cells (Fig. 1g) and is expressed over a much broader region in *as1* extending in multiple directions from the petiole (Fig. 1h). This pattern reflects a change in proximodistal and mediolateral patterning of the leaf.

Loss of *AS1* activity causes defects that are similar, in many respects, to those caused by misexpression of *KNOX* transgenes¹. We used polymerase chain reaction after reverse transcription of RNA (RT-PCR) to examine expression of the *KNOX* genes *STM*, *KNAT1* and *KNAT2* in *as1* mutants. *STM* expression in the wild type is confined to SAM cells and is absent from leaves and leaf primordia¹⁷. No change in the pattern of expression was detected in *as1* (Fig. 2) even at the level of *in situ* hybridization (data not shown). *KNAT1* transcripts in wild-type plants were detected in whole-shoot and inflorescence tissue but not in leaves (Fig. 2), consistent with the reported expression pattern¹⁸. However, *KNAT1* was ectopically expressed in *as1* leaves (Fig. 2) and, as shown by *in situ* hybridization, was ectopically expressed in the cotyledons of *as1* embryos consistent with their altered morphology (Fig. 3l and m). *KNAT2* transcripts were present at high levels in wild-type shoot and inflorescence tissue as expected¹⁹, but also at a low level in wild-type leaves (Fig. 2). In *as1* mutant leaves, expression of *KNAT2* was upregulated (Fig. 2).

We used positional cloning to identify the *AS1* gene. *as1* had previously been mapped to chromosome 2 (ref. 15). Phenotypic and molecular markers were used to map *as1* within a 1.0-centi-Morgans (cM) interval proximal to *det2*. The *Arabidopsis* Genome Initiative sequence in this region²⁰ revealed a likely candidate for

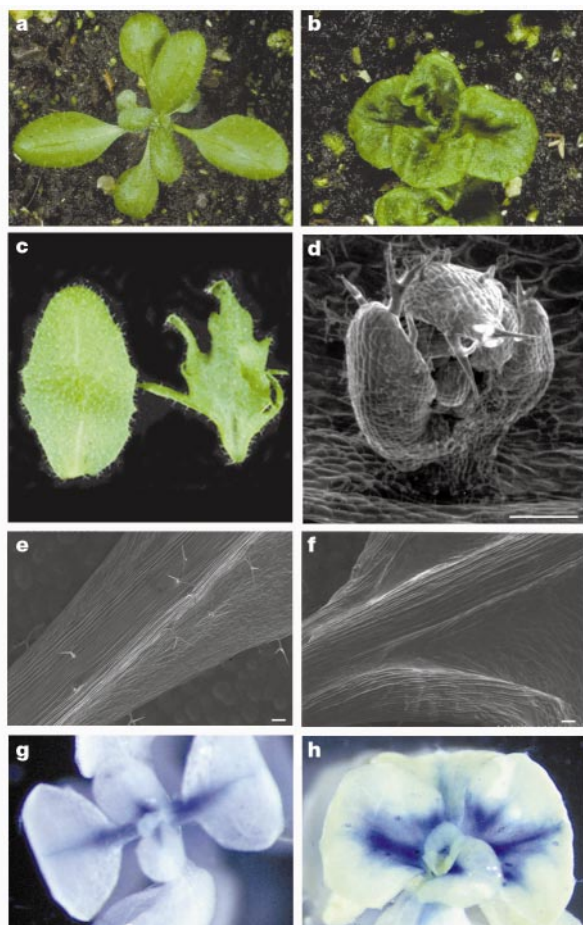


Figure 1 Comparison of *as1* mutant and wild-type *Arabidopsis*. **a**, Vegetative rosette of wild-type. Rosette leaves are elongate and spatulate in shape. **b**, Vegetative rosette of *as1*. Leaves are shorter than wild-type and have a rumpled appearance. **c**, Cauline leaf of the wild type (left), *as1* (right). The *as1* leaf has marginal lobes. **d**, Ectopic shoot in *magnifica*. **e**, Abaxial surface of wild-type leaf at the distal end of the petiole, showing elongated cells of the midvein. **f**, Abaxial surface of *as1* leaf. Elongated epidermal cells typical of the midvein occur in multiple files. **g**, ET2689 in the wild type marks cells of the midvein. **h**, In *as1* ET2689 shows much broadened GUS expression. Scale bars, 200 μm.

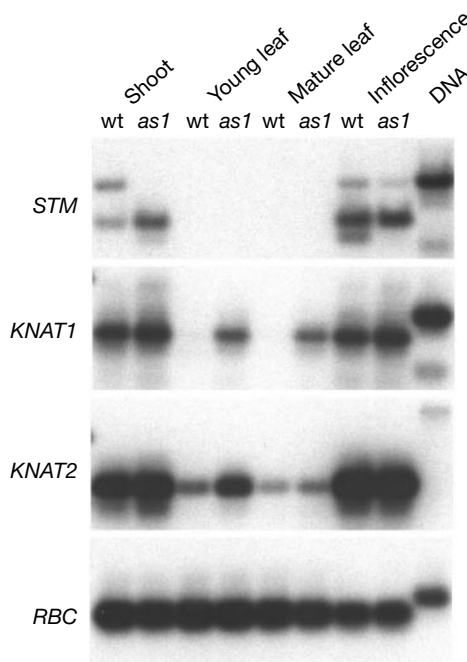


Figure 2 Polymerase chain reaction after reverse transcription of RNA (RT-PCR) analysis of *KNOX* gene expression in *as1* and wild type. RNA was extracted from shoots, young leaves, mature leaves and inflorescence tissues. RT-PCR reactions were performed using gene-specific primers. Products were blotted and hybridized with gene-specific probes. *STM* is detected in wild type (wt) and *as1* shoot and inflorescence, and not in leaf tissue. *KNAT1* is detected in wild type and *as1* shoot and inflorescence and is found in *as1* leaf tissue. *KNAT2* is detected in all tissues in wild type and *as1* but is upregulated in young leaves and, to a lesser extent, in mature leaves of *as1*. *RBC* transcripts were amplified as a control.

AS1 in a gene closely related to the myb transcription factor *PHAN*, which had previously been named *Atphan*¹¹. Two independent *as1* alleles were sequenced and found to have mutations in *Atphan* predicted to truncate the carboxy-terminal third of the *AS1* protein. *as1-magnifica* was found to result from a rearrangement at the 5' end of the gene (likely inversion) that disrupts the promoter. A 6-kilobase (kb) genomic fragment encompassing *Atphan* complemented *as1-1* (see Methods). Together, these results confirmed that *Atphan* is *AS1*.

The expression pattern of *AS1* was examined by *in situ* hybridization. In wild-type embryos *AS1* RNA first became detectable in late globular stage, predominantly in two subepidermal domains corresponding to cotyledon initials (Fig. 3a). Expression was maintained in subepidermal cells of developing cotyledons from heart stage onwards, but was absent or reduced in cells that would subsequently form the shoot apical meristem and the cotyledon epidermis (Fig. 3b–e). After germination, *AS1* was detected in leaf founder cells from the time of primordium initiation until stage P4 and in cells associated with the cotyledon vasculature (Fig. 3f). On flowering, *AS1* RNA was detected on the flank of the inflorescence apex in a region corresponding to the “cryptic bract” (Fig. 3g)²¹. In early floral primordia *AS1* expression was first detected in the abaxial sepal primordium (Fig. 3g), and subsequently in primordia of all floral organs (Fig. 3h). This is consistent with reduced growth of these organs in *as1* flowers¹⁵.

Our analysis of *KNOX* gene expression in *as1* showed misexpression of *KNAT1* and *KNAT2*, whereas *STM* expression was unchanged. To place *STM* in a genetic pathway relative to *AS1*, double mutants were made using a strong allele of *stm*. Embryos homozygous for *stm-1* lack a SAM and develop cotyledons that are fused at their base (Fig. 4b)^{4,5}. F2 progeny from a cross between

stm-1/+ × *as1-1/as1-1* segregated, at a ratio of 1/16, a novel phenotype identified as the *as1-1 stm-1* double mutant (see Methods). Double-mutant vegetative shoots and leaves were indistinguishable from *as1* single mutants (Fig. 4a and c), indicating that the embryonic and vegetative phenotype of *stm-1* was completely suppressed by loss of *AS1* activity. Epistasis of *as1* also demonstrated that the *as1* mutant phenotype was not dependent on *STM* activity, consistent with RT-PCR data showing that these plants do not misexpress *STM*.

In reproductive development, *as1-1 stm-1* double mutants differed from *as1* single mutants in continuing to generate lateral shoots, each subtended by a cauline leaf, instead of flowers (Fig. 4c). The overall shoot architecture was comparable to wild-type plants, although occasional flowers that formed resembled those conditioned by weak alleles of *stm* (Fig. 4d)^{5,22}. The double-mutant phenotype showed that loss of *AS1* activity in *as1-1* was insufficient to suppress the *stm* mutant phenotype in reproductive development.

Genetic interaction between *as1* and *stm* demonstrates that *as1* can rescue the *stm* phenotype in embryonic and vegetative meristems. Given that the two genes are expressed in complementary domains, and that the expression pattern of *STM* does not change in *as1* mutants, we conclude that *STM* negatively regulates *AS1*. *In situ* hybridization of siliques (fruits) segregating 1/4 *stm* embryos revealed *AS1* expression throughout the apical half of 13 out of 60 heart stage embryos consistent with this conclusion (Fig. 3l–k).

The genetic and molecular interaction between *AS1* and *KNOX* genes allows us to propose a mechanism by which organ founder cells are distinguished from stem cells and their derivatives in the shoot apical meristem. In stem cells *STM* negatively regulates *AS1*.

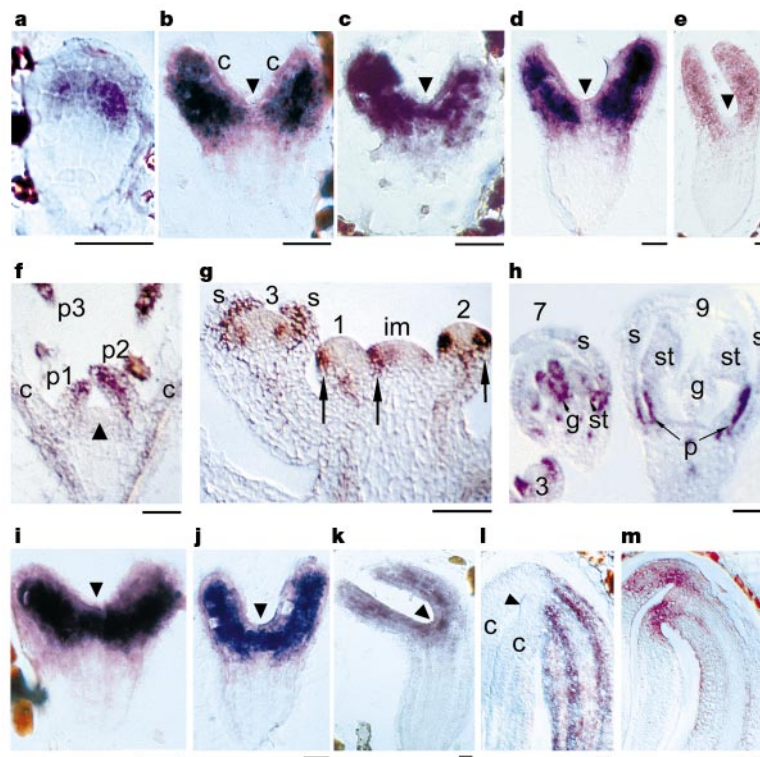


Figure 3 Expression of *AS1* and *KNAT1*. *In situ* hybridization showing expression of *AS1* (a–k) and *KNAT1* (l, m) in wild-type embryos at successive stages of development (a–e), wild-type vegetative apex (f) and inflorescence apices (g, h), *stm-1* mutant (i–k), wild type (l) and *as1* (m) embryos. b and c are two sections from the same embryo taken 7 μm apart. Arrowheads indicate SAM cells or their initials. Arrows in g indicate cryptic bract

and initials of the lowermost sepal. *AS1* expression is confined to organ primordia and is absent from the SAM. In *as1* mutants *KNAT1* is ectopically expressed at the base of the cotyledons. c, cotyledon; im, inflorescence meristem; s, sepal; p, petal; st, stamen; g, gynoecium; p1–p3, leaves of increasing developmental age. Scale bars, 25 μm.

In founder cells, *STM* is downregulated, allowing *AS1* to be expressed. *AS1* in turn downregulates *KNAT1* and *KNAT2*. Thus, *stm* mutant embryos fail to develop a meristem owing to misexpression of *AS1* in presumptive stem cells and their immediate derivatives, causing them to lose their undifferentiated meristematic state. In the *as1 stm-1* double mutant, lack of *AS1* allows meristem function in the absence of *STM*, potentially promoted by derepression of *KNAT1* and *KNAT2*. These *KNOX* genes are closely related to *STM*⁷, although their ability to complement the *stm* phenotype is not known. In floral promordia *STM* has a function independent of *AS1*. The network of negative gene interactions between *AS1* and *KNOX* genes functions to distinguish between stem cells and founder cells within the SAM. This leads to a situation in which *STM* is not required for SAM formation in the absence of *AS1*. In contrast, several other genes known to interact with *STM*, including the *CLV* genes, *WUS* and *CUC2* (refs 23, 24, 25), are all expressed in regions of the SAM and are required for meristem function. Mutations in *CLV1*, *CLV3* and *WUS* show additive interactions with *as1*, demonstrating that *AS1* acts independently of these genes (not shown). This is consistent with *AS1* and *STM* acting in a parallel pathway to *WUS* and *CLV* to specify cell fate in the meristem^{5,22}.

AS1, *PHAN* and *RS2* all negatively regulate *KNOX* gene expression in organ primordia, demonstrating a degree of functional

conservation in *Arabidopsis*, *Antirrhinum* and maize. In maize, *rs2* mutants show alterations along the leaf proximodistal axis, with proximal sheath and ligule tissues displaced distally into the leaf blade¹⁴. *as1* mutants show changes in the proximodistal and mediolateral axes, resulting in characteristic leaf asymmetry (Fig. 1). In *Antirrhinum*, lower leaves of *phan* mutants are broad and heart-shaped with prominent veins similar to those of *as1*. *phan* in addition affects the dorsoventral and proximodistal axes in a cold-sensitive fashion^{10,12}. Upper *phan* leaves are radial and ventralized, and fail to grow out altogether at restrictive temperatures. In contrast, *as1* and *rs2* do not form radialized leaves. *Arabidopsis* and *Antirrhinum* are both eudicots with simple leaves, and fundamental aspects of leaf development are expected to be similar in the two species. The differences between *as1* and *phan* phenotypes may reflect allelic differences or functional divergence of these genes and their targets. Alternatively, they might be explained by modifiers in *Antirrhinum* that enhance *phan* in upper leaves. □

Methods

Plant stocks and growth conditions

All alleles of *as1* (*magnifica*, *as1-1* and *as1-17*) and *stm* used in this study were obtained from the Arabidopsis Biological Resource Center (ARBC). Gene and enhancer trap lines were generated as previously described²⁶. Plants were grown either on soil or on Murashige and Skoog salts media²⁶, supplemented with vitamins, with a minimum day length of 16 h. GUS staining was carried out as previously described²⁶.

Genetic analysis

as1 was mapped relative to *det2* by progeny testing the F2 plants from the cross *cp er as1 cer8* × *det2*. Out of 100 recombinants between *as1* and *cer8*, three carried the haplotype *cp er as1 det2*, placing *as1* 0.3 cM proximal to *det2*. A second cross between *hy1 as1* and ecotype Landsberg *erecta* identified recombinants proximal to *as1*. Molecular mapping of recombinants between *hy1* and *as1* was carried out using derived (this study) and previously identified CAPS markers (ARBC)²⁷. To construct double mutants, plants homozygous for *as1-1* were crossed as males to plants heterozygous for *stm-1*. Double *as1-1 stm-1* mutants segregated in the F2 progeny in the expected 1:15 ratio; percentage segregation for each phenotypic class were wild-type (55.8%): *as1* (17.4%): *stm-1* (21.2%): *as1 stm-1* (5.6%). The *stm-1* genotype of double mutants was confirmed by a PCR assay. For each DNA template a common primer GCCATCATGACATCACATC was used in separate reactions with a primer, CTTTAAGCTCTCTATCCTCAGCTTG, designed to amplify the wild-type *STM* allele and a primer, CTTTAAGCTCTCTATCCTC AGCTTA, designed to amplify the mutant *stm-1* allele¹⁷. Standard PCR reaction conditions were used with an annealing temperature of 66 °C.

DNA and RNA analysis

DNA extraction and manipulation were as described previously²⁸. Total RNA was purified using Trizol reagent (GibcoBRL). Following DNase treatment (Boehringer Mannheim) complementary DNA was synthesized using M-MuLV reverse transcriptase (New England Biolab) in 50 mM Tris-HCl (pH 8.3), 30 mM KCl, 8 mM MgCl₂, 10 mM DTT, 1 mM each of dATP, dCTP, dGTP, dTTP, 1 μM oligo dT, 50 units RNasin, 0.1 μg BSA and 100 units reverse transcriptase. RT-PCR reactions were performed with gene specific primers and products were subject to Southern hybridization using specific probes amplified or subcloned from each gene.

Cloning and sequencing

For complementation a 6-kb genomic fragment encompassing the *AS1* (At2g37630) locus was cloned into the vector pPZP112 (ref. 29) and transformed by *Agrobacterium* into plants. To sequence *as1-1* and *as1-17* alleles genomic DNA from homozygous plants was amplified with the primers ACATTGGAGACACCAATGA and CCCTGTTTGGTTT-CCAGAATA, encompassing the coding region of *AS1*. PCR products were sequenced with internal primers, using the dye terminator cycle sequencing (Applied Biosystems). TAIL-PCR³⁰ was used to determine the sequence disruption in the *magnifica* allele.

Scanning electron microscopy

Fresh material was mounted on silver tape (Electron Microscope Sciences) and viewed with an Hitachi S-3500N SEM using a beam voltage of 5 kV.

In situ hybridization

AS1 RNA was detected with a digoxigenin-labelled riboprobe complementary to part of the *AS1* transcript downstream of the conserved MYB domain (C-terminal 124 amino acids) using the method of ref. 21 (<http://www.wisc.edu/genetics/CATG/barton/protocols.html>).

Received 4 June; accepted 23 October 2000.

1. Martienssen, R. & Dolan, L. in *Arabidopsis. Annual Plant Reviews* (eds Anderson, M. & Roberts, J.) 262–297 (Sheffield Academic Press, Sheffield, 1998).



Figure 4 *as1* shows genetic interaction with *stm-1*. **a**, *as1* whole-plant phenotype. Rosette and cauline leaves are rounded and lobed. After 3–5 cauline leaves with associated secondary shoots, flowers form on the main inflorescence and lateral shoots. **b**, *stm-1* mutants showing cotyledons fused at the base and no vegetative shoot. **c**, *as1 stm-1* double mutant. Rosette is indistinguishable from *as1* single mutants. The main inflorescence and lateral shoots fail to produce flowers, but continue to form lateral shoots with subtending cauline leaves. **d**, *as1 stm-1* double mutant flower. Occasional flowers formed on double mutants typically have reduced petal and stamen number and no central carpel. Floral organs are sometimes mosaic (arrow).

2. Hudson, A. Development of symmetry in plants. *Annu. Rev. Plant. Physiol. Mol. Biol.* **51**, 349–370 (2000).
3. Jackson, D., Veit, B. & Hake, S. Expression of maize *KNOTTED1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. *Development* **120**, 405–413 (1994).
4. Barton, M. K. & Poethig, R. S. Formation of the shoot apical meristem in *Arabidopsis thaliana*—an analysis of development in the wild type and in the shoot meristemless mutant. *Development* **119**, 823–831 (1993).
5. Clark, S. E., Jacobsen, S. E., Levin, J. Z. & Meyerowitz, E. M. The *CLAVATA* and *SHOOT MERISTEMLESS* loci competitively regulate meristem activity in *Arabidopsis*. *Development* **122**, 1567–1575 (1996).
6. Vollbrecht, E., Reiser, L. & Hake, S. Shoot meristem size is dependent on inbred background and presence of the maize homeobox gene, *knotted1*. *Development* **127**, 3161–3172 (2000).
7. Reiser, L., Sanchez-Baracaldo, P. & Hake, S. Knots in the family tree: evolutionary relationships and functions of *knox* homeobox genes. *Plant Mol. Biol.* **42**, 151–166 (2000).
8. Sinha, N. R., Williams, R. E. & Hake, S. Overexpression of the maize homeobox gene, *KNOTTED-1*, causes a switch from determinate to indeterminate cell fates. *Genes Dev.* **7**, 787–795 (1993).
9. Chuck, G., Lincoln, C. & Hake, S. *Kn1* induces lobed leaves with ectopic meristems when overexpressed in *Arabidopsis*. *Plant Cell* **8**, 1277–1289 (1996).
10. Waites, R., Selvadurai, H. R., Oliver, I. R. & Hudson, A. The *PHANTASTICA* gene encodes a MYB transcription factor involved in growth and dorsoventrality of lateral organs in *Antirrhinum*. *Cell* **93**, 779–789 (1998).
11. Timmermans, M. C., Hudson, A., Becraft, P. W. & Nelson, T. ROUGH SHEATH2: a Myb protein that represses *knox* homeobox genes in maize lateral organ primordia. *Science* **284**, 151–153 (1999).
12. Waites, R. & Hudson, A. *phantastica*: a gene required for dorsoventrality of leaves in *Antirrhinum majus*. *Development* **121**, 2143–2154 (1995).
13. Tsiantis, M., Schneeberger, R., Golz, J. F., Freeling, M. & Langdale, J. A. The maize *rough sheath2* gene and leaf development programs in monocot and dicot plants. *Science* **284**, 154–156 (1999).
14. Schneeberger, R., Tsiantis, M., Freeling, M. & Langdale, J. A. The *rough sheath2* gene negatively regulates homeobox gene expression during maize leaf development. *Development* **125**, 2857–2865 (1998).
15. Reidei, G. P. Non-mendelian megagametogenesis in *Arabidopsis*. *Genetics* **51**, 857–872 (1965).
16. Reinholz, E. *Arabidopsis thaliana* (L.) HEYNH. als Objekt für genetische und entwicklungsphysiologische Untersuchungen. *Arabidopsis Inform. Service* **01S**, 24 (1965).
17. Long, J. A., Moan, E. L., Medford, J. I. & Barton, M. K. A member of the *KNOTTED* class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. *Nature* **379**, 66–69 (1996).
18. Lincoln, C., Long, J., Yamaguchi, J., Serikawa, K. & Hake, S. A *knotted1*-like homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. *Plant Cell* **6**, 1859–1876 (1994).
19. Dockx, J. *et al.* The homeobox gene *ATK1* of *Arabidopsis thaliana* is expressed in the shoot apex of the seedling and in flowers and inflorescence stems of mature plants. *Plant Mol. Biol.* **28**, 723–737 (1995).
20. Lin, X. *et al.* Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **402**, 761–768 (1999).
21. Long, J. & Barton, M. K. Initiation of axillary and floral meristems in *Arabidopsis*. *Dev. Biol.* **218**, 341–353 (2000).
22. Endrizzi, K., Moussian, B., Haecker, A., Levin, J. Z. & Laux, T. The *SHOOT MERISTEMLESS* gene is required for maintenance of undifferentiated cells in *Arabidopsis* shoot and floral meristems and acts at a different regulatory level than the meristem genes *WUSCHEL* and *ZWILLE*. *Plant J.* **10**, 101–113 (1996).
23. Clark, S. E., Running, M. P. & Meyerowitz, E. M. *CLAVATA3* is a specific regulator of shoot and floral meristem development affecting the same processes as *CLAVATA1*. *Development* **121**, 2057–2067 (1995).
24. Mayer, K. F. *et al.* Role of *WUSCHEL* in regulating stem cell fate in the *Arabidopsis* shoot meristem. *Cell* **95**, 805–815 (1998).
25. Aida, M., Ishida, T. & Tasaka, M. Shoot apical meristem and cotyledon formation during *Arabidopsis* embryogenesis: interaction among the *CUP-SHAPED COTYLEDON* and *SHOOT MERISTEMLESS* genes. *Development* **126**, 1563–1570 (1999).
26. Sundaresan, V. *et al.* Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Gene Dev.* **9**, 1797–1810 (1995).
27. Li, J., Nagpal, P., Vitart, V., McMorris, T. C. & Chory, J. A role for brassinosteroids in light-dependent development of *Arabidopsis*. *Science* **272**, 398–401 (1996).
28. Springer, P. S., McCombie, W. R., Sundaresan, V. & Martienssen, R. A. Gene trap tagging of *PROLIFERA*, an essential MCM2-3-5-like gene in *Arabidopsis*. *Science* **268**, 877–880 (1995).
29. Hajdukiewicz, P., Svab, Z. & Maliga, P. The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. *Plant Mol. Biol.* **25**, 989–994 (1994).
30. Liu, Y. G., Mitsukawa, N., Oosumi, T. & Whittier, R. F. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* **8**, 457–463 (1995).

Acknowledgements

We thank A. Groover, C. Kidner, E. Vollbrecht, M. Timmermans, J. Golz and D. Jackson for helpful discussions, Q. Gu, P. Springer, J. Li and J. Chory for help with mapping and T. Laux for wus-1 seed. We also thank T. Mulligan for plant care, and K. Schutz and D. McCombie for help with sequencing. This work was supported by a Human Frontiers Science Program postdoctoral fellowship to M.C., a Biotechnology and Biological Sciences Research Council studentship to R.B. and grant support from the National Science Foundation, Department of Energy and United States Department of Agriculture to R.M.

Correspondence and requests for materials should be addressed to R.M. (e-mail: martiens@cshl.org).

Neuronal switching of sensorimotor transformations for antisaccades

Mingsha Zhang & Shabtai Barash

Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel

The influence of cognitive context on orienting behaviour can be explored using the mixed memory-prosaccade, memory-antisaccade task. A symbolic cue, such as the colour of a visual stimulus, instructs the subject to make a brief, rapid eye movement (a saccade) either towards the stimulus (prosaccade) or in the opposite direction (antisaccade)^{1–3}. Thus, the appropriate sensorimotor transformation must be switched on to execute the instructed task. Despite advances in our understanding of the neuronal processing of antisaccades^{4–8}, it remains unclear how the brain selects and computes the sensorimotor transformation leading to an antisaccade. Here we show that area LIP of the posterior parietal cortex is involved in these processes. LIP's population activity turns from the visual direction to the motor direction during memory-antisaccade trials. About one-third of the visual neurons in LIP produce a brisk, transient discharge in certain memory-antisaccade trials. We call this discharge 'paradoxical' because its timing is visual-like but its direction is motor. The paradoxical discharge shows, first, that switching occurs already at the level of visual cells, as previously proposed by Schlag-Rey and colleagues⁹; and second, that this switching is accomplished very rapidly, within 50 ms from the arrival of the visual signals in LIP.

Figure 1a illustrates the problem raised in ref. 5. At the core of the sensorimotor transformation guiding antisaccades lies vector inversion. How does the brain compute this inversion? Figure 1b sketches relevant processing at the single-cell level. Figure 1b refers to visual and motor activity; a standard criterion for distinguishing between visual and motor activity is by timing the activity in memory-saccades^{9–15}, as shown below. For brevity, we call cells with visual activity 'visual cells', and cells with motor activity 'motor cells'. Although Fig. 1b is simplistic, it provides a framework for the two hypotheses we will consider for task switching. Both hypotheses propose that prosaccades and antisaccades involve alternative sets of functional connections between cells. A standard set of connections leads to prosaccades. A different set of connections, that emerges during training, leads to antisaccades. An instruction to perform an antisaccade switches the set of active connections from one configuration to the other.

According to the "visual switching hypothesis" (proposed in ref. 5), the switched connections are between visual cells. The process of switching connections would be reflected in a critical set of visual cells, each having two alternative receptive fields, a standard receptive field for ordinary prosaccades and another receptive field in the opposite direction specifically for antisaccades. Activation of the new receptive field, through the new connections, would instigate vector inversion. Thus, vector inversion would be accomplished by receptive field remapping¹⁶. The "visuomotor switching hypothesis" proposes that switching acts on connections between visual and motor cells. The two hypotheses lead to contrasting experimental predictions with regard to each cell's visual receptive field (mapped relative to the visual vector) and motor field (mapped relative to the motor vector) in prosaccades and antisaccades. The visuomotor switching hypothesis predicts that both visual and motor fields remain unchanged, in all cells. Although a critical set of cells with motor activity would fire in trials with opposite visual vectors in prosaccades and antisaccades, the motor fields of these cells would not change. In contrast, the visual switching hypothesis predicts changed receptive fields for some

visual cells. Therefore, at the outset of this study we considered the visuomotor hypothesis more likely.

We recorded neuronal activity from area LIP in two monkeys trained in memory-antisaccades and related tasks. When we isolated a cell, we determined its preferred direction by testing the cell's activity in memory-prosaccades made in 12 directions. We then studied the cell's activity with a mixed memory-prosaccade memory-antisaccade task (Fig. 1c). Because the instruction to make a prosaccade or an antisaccade is specified by the colour of the stimulus, which varies from trial to trial, the monkey can determine the movement direction appropriate for the current trial only upon stimulus presentation. Therefore, all sensorimotor transformation switching must follow stimulus onset.

The stimulus appeared in either of two opposite locations, always positioned so that one location fell in the cell's preferred direction and the other in the opposite direction. Thus, there were four types of trials, two prosaccades (V+M+ and V-M-) and two antisaccades (V+M- and V-M+). Here V+ and M+ signify that the visual stimulus, or the movement, respectively, are in the preferred direction, and V- and M- mean that they are opposite to the preferred direction⁶.

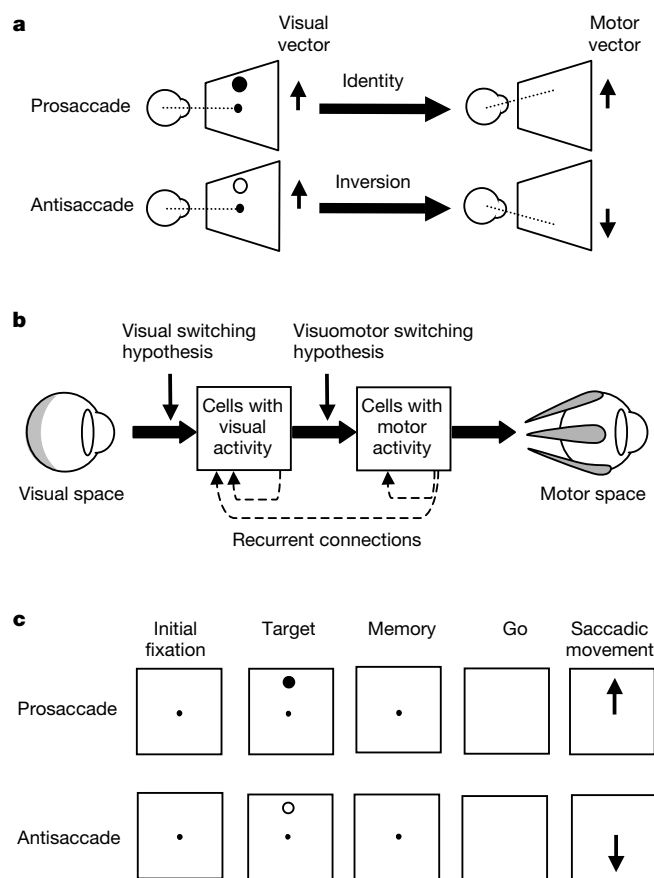


Figure 1 Cartoons of involved operations, hypotheses and task details. **a**, The sensorimotor transformations necessary for prosaccades and for antisaccades. The visual vector is defined as originating in the central fovea and terminating at the image of the visual stimulus. **b**, A cartoon showing sites where antisaccade-switching signals could act on the flow of information mediating ordinary saccades. Arrows should not be taken as literal direct connections; intermediates between visual and motor responses and multiple brain regions are involved. **c**, The task. The two types of trials differ only in the colour of the stimulus and the required movement (red stimulus signals a request for prosaccade, green for antisaccade). Typical durations are presaccadic fixation, 500 ms; stimulus, 250 ms; memory, 1,000 ms; postsaccadic fixation, 250 ms. To maintain movement precision in these conditions, at the end of some trials the stimulus reappeared and the monkey had to make a visually guided saccade to its location.

Figure 2a illustrates the activity of an LIP visual neuron. A brisk discharge briefly follows stimulus presentation in V+ trials, regardless of the subsequent saccade's direction (both V+M+ and V+M-). The discharge ends long before the movement. Therefore, this discharge is indeed visual, and is not motor. Moreover, the neuron remains almost inactive when visual stimuli are presented in the opposite direction (V- trials), regardless of saccade direction (V-M+ and V-M-). The activity of the cell illustrated in Fig. 2b is closely linked to the movement's planning and execution; for brevity we refer to such neurons as motor cells. Long after stimulus offset, a discharge builds up in M+ trials, gradually increasing towards the time of the saccadic movement (V+M+ and V-M+). The cell remains inactive in M- trials (V+M- and V-M-); hence this discharge is movement-related, or motor.

We computed for each neuron a 'differential activity' (DA) vector, mapping the neuron's activity in antisaccades onto the range between 'visual' and 'motor'. We define $DA(t) = R_{V-M+}(t) - R_{V+M-}(t)$ where $R_A(t)$ denotes the mean spike rate of the given neuron in trial group *A* at time *t* in the trial. We estimate the DA of a neuron as the bin-by-bin difference between the time-histograms of the two types of antisaccade trials, V-M+ and V+M-. Consistent with this definition, the visual neuron's DA is negative

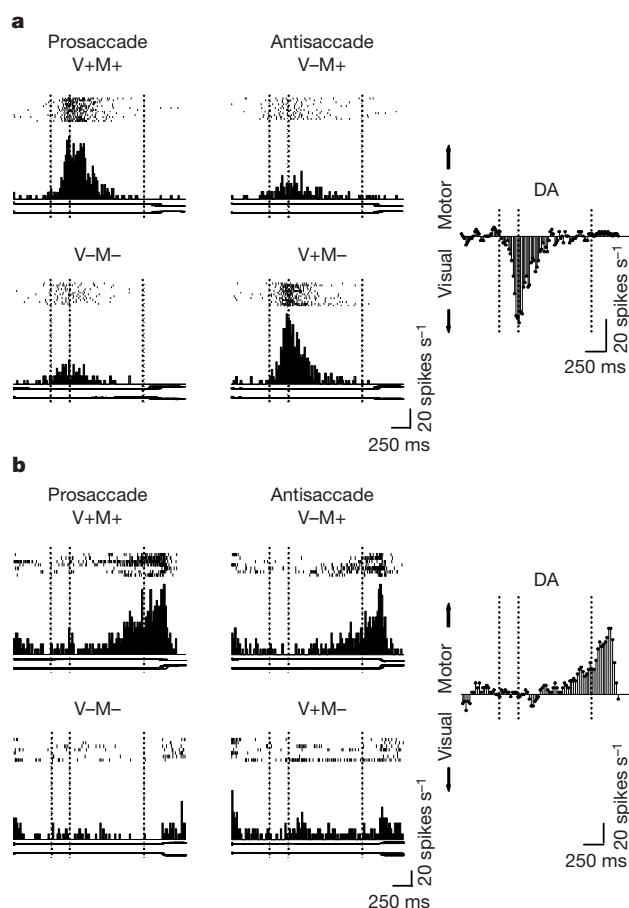


Figure 2 Responses of single PPC neurons. **a**, A visual cell and **b**, a motor cell. The four panels on the left of each part of the figure illustrate the neuron's response in trials of the types listed in the panel headers. Each panel contains, from the top, a raster illustrating spike timings, with each row showing a separate trial; a spike histogram (bin width 20 ms); and vertical and horizontal eye position traces. The panel on the right shows the differential activity vector (DA) of each cell. Time axis is the same for all panels. The scale of all four histograms of each part is shown on the left. Dashed vertical lines mark, from the left, onset and offset times of the stimulus and offset of the fixation spot (the 'go' signal). Thus, the stimulus is presented between the first two dashed vertical lines; the memory interval takes place between the second and third vertical lines.

throughout the trial (Fig. 2a) and the motor neuron's DA is positive throughout (Fig. 2b). We therefore call the negative direction of the DA 'visual' and the positive direction 'motor' (see Methods).

Figure 3 shows the mean DA of the full sample of 400 units that we studied, together with its pointwise 95% confidence interval. Soon after the onset of the visual stimulus (latency about 55 ms), the population DA turns towards the visual direction. The visual phase is brisk but brief: the peak of the visual phase is reached within another 100 ms. Afterwards the population DA rapidly declines and crosses over towards the motor direction. The population DA then gradually increases in the motor direction until the time of the saccadic movement, and only then goes down. Thus, the population DA has two clearly segregated phases, an early visual phase followed by a motor phase. The inversion from visual to motor occurs in the middle of the memory interval: by its pointwise 95% confidence interval, the zero crossing of the DA is between 140 and 690 ms (mean 415 ms) into the 1-s memory interval. The population DA thus provides a neuronal correlate of the transformation underlying antisaccade performance.

What events on the single-cell level trigger the inversion of the population activity? A novel type of discharge is illustrated by another LIP neuron (Fig. 4a). Three of the four conditions in Fig. 4a are consistent with the interpretation that this neuron is visual. Indeed, a visual response is evident in both V+ conditions (V+M+ and V+M-), as compared to the baseline condition (V-M-). However, this cell discharges vigorously also in the non-visual condition, V-M+. This discharge is paradoxical: although it is not visual, it appears to be visual in its time course. Its latency is brief, well within the range of visual response latencies¹²; and it declines to baseline during the memory period, long before the movement. Thus, the paradoxical activity differs markedly in its

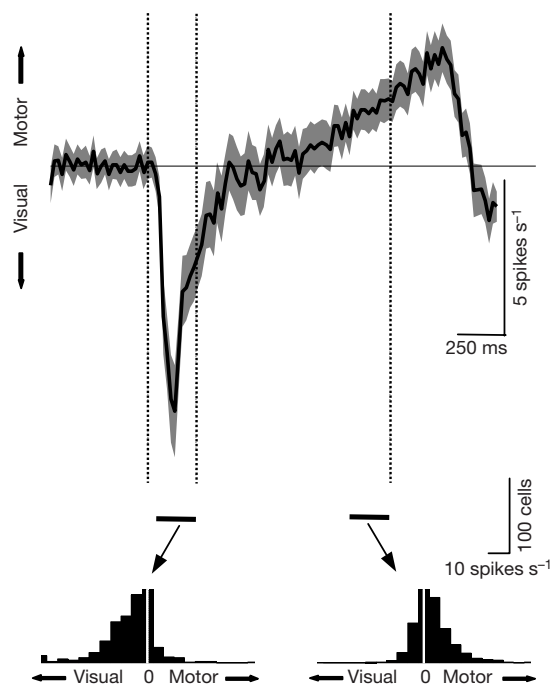


Figure 3 Mean differential activity for all 400 units studied. (Summed DA divided by number of units.) The DA is plotted in black. The pointwise 95% confidence interval, computed for each 20-ms bin, is shown in grey. Dashed vertical lines show, from the left, times of stimulus onset, stimulus offset and fixation spot offset, as in Fig. 2. The insets at the bottom show histograms of the mean DA computed for the two displayed 200-ms intervals, for each of the 400 units. The intervals span the visual response (50 to 250 ms after stimulus onset) and the last 200 ms of the memory interval. The histograms are clipped at 100 cells; actual values of the bins at 0 spikes per second are 132 and 163 cells for the left and right histograms, respectively.

time-course from the build-up activity of visuomotor and motor cells (Fig. 2b).

We analysed a set of 185 units selected to have strong visual response and little or no motor activity in memory-prosaccades, regardless of their activity in memory-antisaccades. This set

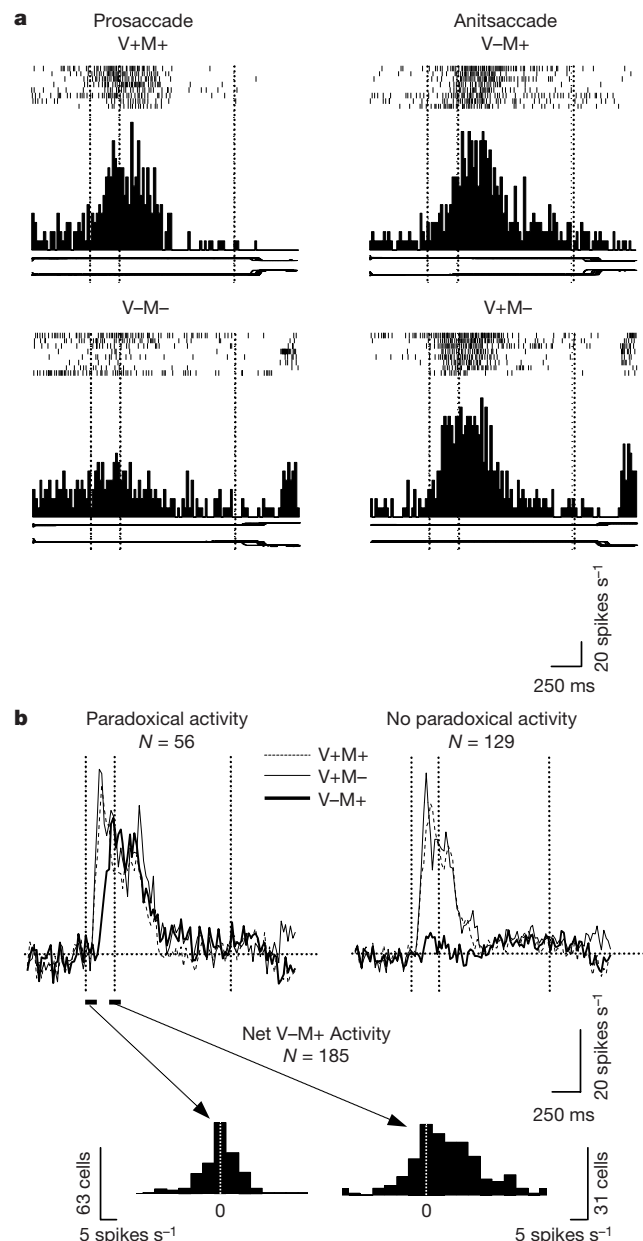


Figure 4 Illustration of the paradoxical activity. **a**, Activity of a visual cell with paradoxical activity, illustrated in the same format as in Fig. 2. **b**, Mean net responses in the paradoxical condition compared with the mean visual responses, for two complementary subsets of a set of 185 cells with visual activity but little or no motor activity. A net response is defined as the mean activity in the relevant condition minus the mean activity in the V-M-, baseline condition. The mean V-M+ response distinctly shows the paradoxical activity in the first group ($N = 56$, 30%) but is flat in the second group ($N = 129$, 70%). Vertical dashed lines have the same meaning as in Fig. 2. The insets at the bottom show histograms of the mean net activity in the V-M+ condition computed over 100-ms intervals before (left inset, 0 to 100 ms after stimulus onset) and during the paradoxical activity (right inset, 200 ms to 300 ms after stimulus onset). The histograms pull together data from both subsets displayed above. The mean of the left histogram is not significantly different from zero ($P = 0.78$); the mean of the second histogram, 5.53 spikes per second, is very significantly different from zero ($P = 0$ using standard double precision).

could be divided into two subsets, with ($N = 56, 30\%$) and without ($N = 129, 70\%$) substantial paradoxical activity. Figure 4b shows the mean net responses of these two groups. The net visual response in memory-prosaccades, $R_{V+M+}(t) - R_{V-M-}(t)$ is depicted as thin broken lines. The net visual response in memory-antisaccades, $R_{V+M-}(t) - R_{V-M+}(t)$ is depicted as thin solid lines. The non-visual condition, $R_{V-M+}(t) - R_{V-M-}(t)$ is depicted as thick solid lines. All four mean visual responses (thin lines) are similar. In contrast, the mean activity in the non-visual condition $V-M+$ (thick lines) shows a brisk discharge of paradoxical activity for the first group of cells, but remains at baseline for the second group. The latency of all four mean visual responses is about 60 ms. The latency of the mean paradoxical response is about 110 ms. Therefore, within 50 ms from the time of arrival of visual responses LIP is activated by input contingent on the task switch.

The presence of visual cells with paradoxical activity is consistent with the visual switching hypothesis (Fig. 1b). The paradoxical discharge might represent a remapped visual response to the oppositely directed stimulus, activated by nonstandard visual input connections. This nonstandard input pathway would usually be inactive. During the 110 ms after the onset of the visual stimulus, some context-categorization process switches these input connections on. After the 50-ms delay caused by context-categorization, the mean paradoxical activity increases rapidly, and then overlaps the mean visual response (Fig. 4b). This time course is consistent with the interpretation that the paradoxical activity is a visual response to the stimulus in the remapped receptive field.

Alternatively, the paradoxical activity might represent direct activation of the visual cell through its recurrent connections. According to this alternative interpretation, the cell's receptive field remains fixed. The paradoxical activity would reflect recurrent, top-down activation, rather than visual, bottom-up. Two waves would thus sweep through LIP shortly after stimulus presentation in antisaccade trials. Bottom-up visual responses, making up the first wave, would spread throughout the brain and feed into a context-categorization process. Within another 50 ms, the results of the context-categorization would reach LIP in the form of the second wave, consisting of the top-down paradoxical activity. Further experiments are needed to distinguish between these two interpretations. Furthermore, although the presence of the paradoxical activity shows that visual switching exists, it is still unclear whether visuomotor switching occurs too.

Our findings are consistent with Andersen's proposal in 1987 of posterior parietal cortex involvement in sensorimotor transformations¹⁷. Scalp recordings in humans performing antisaccades support our findings¹⁸. A recent study of antisaccades in LIP⁶ addressed related but different questions and used substantially different procedures. It is possible that the late presaccadic signal described in ref. 6 is a variation of the paradoxical activity. The conclusions of ref. 6 were based mostly on saccades-on-demand; memory-saccades were made only in one direction, precluding direct comparison with our results. Finally, motor-intention dependent visual responses were recently described in the superior colliculus¹⁹; in another study, LIP auditory response fields were acquired by training²⁰. Perisaccadic receptive field remapping¹⁶ in some visual cells was also described in LIP. It remains to be seen whether the paradoxical activity has anything in common with these responses.

Thus, area LIP's population activity reflects the inversion transformation guiding antisaccades. About a third of LIP's visual cells show a novel type of discharge in antisaccade trials that is consistent with the visual switching hypothesis. The paradoxical activity might be a remapped visual response, or alternatively direct activation by recurrent connections. Task switching might also affect visuomotor connections. Altogether, our results reject the visuomotor connections as the sole site of switching, and therefore reject the visuomotor switching hypothesis. □

Methods

Two *Macaca fascicularis* monkeys were implanted with chronic-recording chambers placed over their posterior parietal lobules and scleral search coils for eye-position measurements. All techniques are standard. The experimental procedures conform to the US NIH guidelines and Israeli law and were approved by the Animal Care and Use Committee of the Weizmann Institute. According to standard physiological criteria^{11,12} the cells reported here were from area LIP. Nevertheless, a minority of the cells may have come from area MP²¹ or other neighbouring regions^{17,22}. The monkeys were first trained to perform memory-prosaccades with more than a second of memory, and only then to make memory-antisaccades. The task is illustrated in Fig. 1c. The red and green stimuli are isoluminant disks of the same size. The monkeys were also trained and tested in other oculomotor tasks. Some blocks contained the task of the present paper together with other conditions. We verified that these do not interact with the conditions described here by testing 12 cells with and without the additional conditions; the discharge characteristics remained unchanged. All recordings were conducted with the monkey in a completely darkened, sound-attenuated room. Fuller descriptions of the behavioural methods are given elsewhere²³. Following the isolation of a cell, we first mapped its preferred direction by testing the activity in memory-prosaccades made towards a target in one of 12 equally spaced locations on a 15° circle around a central fixation spot. Typical LIP cells have a clear preferred direction, and in visuomotor cells this direction is common to the visual, memory and motor discharges¹¹. All the cells in the present study had clear preferred directions, and were much less activated, if at all, in memory-prosaccades made in the opposite directions.

The DA represents the population vector

The DA approximates a given cell's net contribution to the population vector for antisaccades: The motor vectors produced in $V+M-$ and $V-M+$ trials are opposite in direction and equal in size. Therefore, the DA is a vector sum of the intended movement vectors weighted by the momentary spike rates: $DA(t) = R_{V+M-}(t)V_{V+M-} + R_{V-M+}(t)V_{V-M+}$, where V_A is the intended-movement vector of class A. This leads to the following interpretation of the DA. The preferred directions of LIP neurons are distributed roughly uniformly in all directions. Let us therefore assume, following ref. 24, that another LIP neuron could be matched to each neuron in our sample so that the temporal profile of the activity of that neuron would be identical to our neuron, but the preferred direction would be opposite. The net effect of the two neurons on the population vector would be $DA(t)$.

The paradoxical activity is not a visual response

The PPC is part of the dorsal visual stream and its neurons are known to be colour insensitive. Stimulus colour was used in previous PPC studies to indicate task context¹⁴. Nevertheless, we conducted a control study and tested the response of 26 visual neurons and another 8 visuomotor neurons in sequences of isoluminant red-blue-green-red, or red-green-green-red stimuli separated by memory intervals. The recordings were conducted long after the monkeys were overtrained in the task of Fig. 1c; the monkeys were easily trained to perform memory-prosaccades in response to these stimulus sequences. The differences between the responses to green and red in the control condition were negligible. All the neurons responded to a green stimulus in the opposite direction in a way similar to their response to a red stimulus in that location, in both cases much more weakly than to a stimulus of either colour in the receptive field. Therefore, the colour difference does not account for the paradoxical activity.

Received 22 May; accepted 19 September 2000.

- Hallett, P. E. Primary and secondary saccades to goals defined by instructions. *Vision Res.* **18**, 1279–1296 (1978).
- Amador, N., Schlag-Rey, M. & Schlag, J. Primate antisaccades. I. Behavioral characteristics. *J. Neurophysiol.* **80**, 1775–1786 (1998).
- Everling, S. & Fischer, B. The antisaccade: a review of basic research and clinical studies. *Neuropsychologia* **36**, 885–899 (1998).
- Funahashi, S., Chafee, M. V. & Goldman-Rakic, P. S. Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. *Nature* **365**, 753–756 (1993).
- Schlag-Rey, M., Amador, N., Sanchez, H. & Schlag, J. Antisaccade performance predicted by neuronal activity in the supplementary eye field. *Nature* **390**, 398–401 (1997).
- Gottlieb, J. & Goldberg, M. E. Activity of neurons in the lateral intraparietal area of the monkey during an antisaccade task. *Nature Neurosci.* **2**, 906–912 (1999).
- Everling, S., Dorris, M. C., Klein, R. M. & Munoz, D. P. Role of primate superior colliculus in preparation and execution of anti-saccades and pro-saccades. *J. Neurosci.* **19**, 2740–2754 (1999).
- Everling, S. & Munoz, D. P. Neuronal correlates for preparatory set associated with pro-saccades and anti-saccades in the primate frontal eye field. *J. Neurosci.* **20**, 387–400 (2000).
- Hikosaka, O. & Wurtz, R. H. Visual and oculomotor functions of monkey substantia nigra pars reticulata. III. Memory-contingent visual and saccade responses. *J. Neurophysiol.* **49**, 1268–1284 (1983).
- Gnadt, J. W. & Andersen, R. A. Memory related motor planning activity in posterior parietal cortex of macaque. *Exp. Brain Res.* **70**, 216–220 (1988).
- Barash, S., Bracewell, R. M., Fogassi, L., Gnadt, J. W. & Andersen, R. A. Saccade-related activity in the lateral intraparietal area. II. Spatial properties. *J. Neurophysiol.* **66**, 1109–1124 (1991).
- Barash, S., Bracewell, R. M., Fogassi, L., Gnadt, J. W. & Andersen, R. A. Saccade-related activity in the lateral intraparietal area. I. Temporal properties; comparison with area 7a. *J. Neurophysiol.* **66**, 1095–1108 (1991).
- Colby, C. L., Duhamel, J. R. & Goldberg, M. E. Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *J. Neurophysiol.* **76**, 2841–2852 (1996).
- Snyder, L. H., Batista, A. P. & Andersen, R. A. Coding of intention in the posterior parietal cortex. *Nature* **386**, 167–170 (1997).

15. Chafee, M. V. & Goldman-Rakic, P. S. Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J. Neurophysiol.* **79**, 2919–2940 (1998).
16. Duhamel, J. R., Colby, C. L. & Goldberg, M. E. The updating of the representation of visual space in parietal cortex by intended eye movements. *Science* **255**, 90–92 (1992).
17. Andersen, R. A. in *Handbook of Physiology, Section 1: The Nervous System* (eds Mountcastle, V. B., Plum, F. & Geiger, S. R.) 483–518 (Am. Physiol. Soc., Bethesda, 1987).
18. Everling, S., Spantekow, A., Krappmann, P. & Flohr, H. Event-related potentials associated with correct and incorrect responses in a cued antisaccade task. *Exp. Brain Res.* **118**, 27–34 (1998).
19. Horwitz, G. D. & Newsome, W. T. Separate signals for target selection and movement specification in the superior colliculus. *Science* **284**, 1158–1161 (1999).
20. Grunewald, A., Linden, J. F. & Andersen, R. A. Responses to auditory stimuli in macaque lateral intraparietal area. I. Effects of training. *J. Neurophysiol.* **82**, 330–342 (1999).
21. Thier, P. & Andersen, R. A. Electrical microstimulation distinguishes distinct saccade-related areas in the posterior parietal cortex. *J. Neurophysiol.* **80**, 1713–1735 (1998).
22. Colby, C. L., Gattass, R., Olson, C. R. & Gross, C. G. Topographical organization of cortical afferents to extrastriate visual area PO in the macaque: a dual tracer study. *J. Comp. Neurol.* **269**, 392–413 (1988).
23. Barash, S., Melikyan, A., Sivakov, A. & Tauber, M. Shift of visual fixation dependent on background illumination. *J. Neurophysiol.* **79**, 2766–2781 (1998).
24. Newsome, W. T., Britten, K. H. & Movshon, J. A. Neuronal correlates of a perceptual decision. *Nature* **341**, 52–54 (1989).

Acknowledgements

We thank E. Ahissar for his involvement, and A. Melikyan and X. Wang for participation in some experiments. We thank M. Glickstein, P. Thier, E. Seidemann, Y. Ritov and M. Tsodyks for discussions and for reading the manuscript. This work was supported by the Israel Science Foundation, and by the Murray H. and Meyer Grodetsky Center for Research of Higher Brain Functions and the Einhorn-Dominic Institute of Brain Research at the Weizmann Institute, and by the Paul Godfrey Research Foundation.

Correspondence and requests for materials should be addressed to S.B. (e-mail: shabtai.barash@weizmann.ac.il).

A learning deficit related to age and β -amyloid plaques in a mouse model of Alzheimer's disease

Guiquan Chen*, Karen S. Chen†, Jane Knox*, Jennifer Inglis*, Andrew Bernard*, Stephen J. Martin*, Alan Justice†, Lisa McConlogue†, Dora Games†, Stephen B. Freedman† & Richard G. M. Morris*

* Department of Neuroscience, University of Edinburgh, 1 George Square, Edinburgh, EH8 9JZ, UK

† Elan Pharmaceuticals, 800 Gateway Boulevard, South San Francisco, California 94080, USA

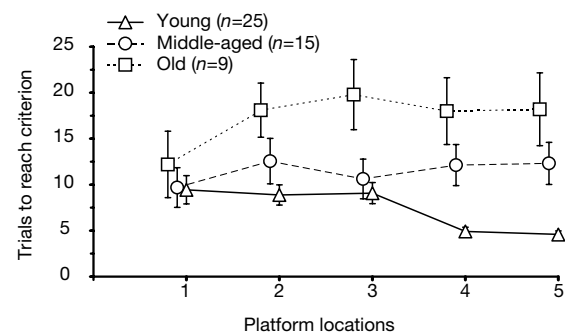
Mice that overexpress the human mutant amyloid precursor protein (hAPP) show learning deficits, but the apparent lack of a relationship between these deficits and the progressive β -amyloid plaque formation that the hAPP mice display is puzzling. In the water maze¹, hAPP mice are impaired before and after amyloid plaque deposition^{2–7}. Here we show, using a new water-maze training protocol, that PDAPP mice⁸ also exhibit a separate age-related deficit in learning a series of spatial locations. This impairment correlates with β -amyloid plaque burden and is shown in both cross-sectional and longitudinal experimental designs. Cued navigation and object-recognition memory are normal. These findings indicate that A β overexpression and/or A β plaques are associated with disturbed cognitive function and, importantly, suggest that some but not all forms of learning and memory are suitable behavioural assays of the progressive cognitive deficits associated with Alzheimer's-disease-type pathologies.

We have developed a new water-maze protocol for mice in which, after they learn to escape quickly and reliably onto the hidden platform at one location, we move the platform to a new location. Numerous locations are used successively, one at a time. In such a procedure, earlier locations of the platform will be encoded in long-term memory, potentially causing interference. Memory retrieval must therefore be selective for the most recently encoded location,

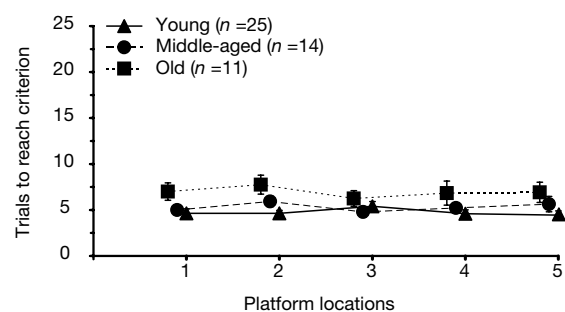
an 'episodic-like' component of the task. The idea behind using such a procedure, embedded within a larger test battery, derives from lesion and single-unit recording studies in rodents that implicate the hippocampal formation in only certain aspects of spatial, working- or episodic-like memory^{9–13}. Such a protocol is worth exploring as, in the PDAPP mouse, the highest density of β -amyloid plaques is first seen in the outer molecular layer of the dentate gyrus and other regions of the hippocampus.

PDAPP and non-transgenic (non-Tg) littermate control mice ($n = 99$) swam normally and climbed successfully onto the escape

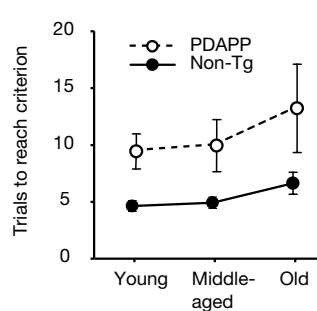
a PDAPP groups



b Non-Tg groups



c Location 1



d Locations 4 and 5

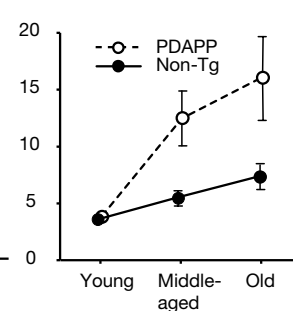


Figure 1 Age-related and age-independent deficits in spatial learning in PDAPP mice. Analysis of variance of 'trials to criterion' (3 successive trials with mean escape latency < 20 s) for each platform location of the first place navigation task revealed a significant groups $\times$ age effect ($F = 5.27$, d.f. 2/93, $P < 0.01$). This interaction justified separate analyses of the two groups. **a**, PDAPP mice. Trials to criterion (inclusive) for each successive platform location in the main series of five locations revealed a significant age-related worsening of performance ($F = 8.32$, d.f. 2/46, $P = 0.001$) and a locations $\times$ age interaction ($F = 2.17$, d.f. 6.59/151.7, $P < 0.05$; Greenhouse–Geisser correction). **b**, Non-Tg mice also showed a modest decline with age ($F = 7.24$, d.f. 2/47, $P < 0.005$) but no interaction. **c**, Platform location 1. A separate analysis of platform location 1 revealed poorer performance by the PDAPP mice which learned more slowly than non-Tgs ($F = 15.2$, d.f. 1/93, $P < 0.001$) but there was no age-related effect or interaction. **d**, Platform locations 4 and 5. There were significant main effects of group and age and, critically, a group $\times$ age interaction ($F = 8.30$, d.f. 2/93, $P < 0.001$). This was because the PDAPP mice got worse with age ($F = 12.56$, d.f. 2/46, $P < 0.0005$) at a much faster rate than non-Tgs ($F = 4.45$, d.f. 2/47, $P < 0.05$). Results shown are mean $\pm$ 1 s.e.m.

platform in the pool. Performance in the initial cued-navigation task (see Methods) revealed no persistent sensorimotor or motivational abnormalities. Escape latencies averaged less than 10 s in all groups by the end of this pretraining.

Two key findings emerged in the new 'training-to-criterion' task, which requires the learning of five successive platform locations. The first of these (cross-sectional study) is a striking age-related deficit in the number of trials required to reach the performance criterion as the mice learned the series of individual platform locations (Fig. 1a). Young PDAPP mice (6–9 months) improved such that criterion was reached in fewer trials for later platform locations, whereas middle-aged (13–15 months) and old PDAPP animals (18–21 months) took longer to learn each new location. The non-Tg control groups did not show this pattern (Fig. 1b). Second, the PDAPP groups of all ages took longer to learn the first platform location of the series than the non-Tg groups (Fig. 1c). This second deficit was age-independent, in contrast to the age-related learning deficit apparent for spatial locations 4 and 5 in the PDAPP groups (Fig. 1d).

Measured over the first three trials of training, swim speeds of PDAPP and non-Tg mice declined in parallel as a function of age, with PDAPP mice swimming only slightly more slowly (0.024 m s^{-1}) than non-Tgs (Fig. 2a, b). Fig 2c shows the swim-paths of representative animals for all trials of location 5 at each age. The young PDAPP mouse reached the performance criterion (trials labelled C1 to C3) as quickly as its non-Tg control (top two rows). In contrast, the middle-aged and old PDAPP mice displayed persistent deficits (taking 13 and 19 trials, respectively, to learn, compared with 4 and 6 trials by non-Tgs). The paths taken by the older PDAPP animals were noteworthy in two respects. First, path length showed great variability from trial to trial until the criterion was reached, suggestive of a retrieval deficit. Second, there was little indication of thigmotaxis (that is, staying near the side walls) which is a pattern of behaviour frequently displayed in the standard reference memory task by mice that fail to learn well.

The test of 'learning capacity' provided a point of comparison between the cross-sectional and longitudinal components of the

study. Both revealed a greater age-related decline in the PDAPP groups (Fig. 3). This deficit was remarkably similar in both the cross-sectional (Fig. 3a) and longitudinal (Fig. 3b) data.

The brains of a subset of PDAPP mice in the cross-sectional study were analysed using the anti-A β monoclonal antibody 3D6 which shows diffuse plaques. Very few plaques were visible in scattered regions of the hippocampal formation at 9 months (Fig. 4a, young mutant animal). By 16–17 months, plaque density was higher, particularly in the termination zones of the axons of layers II and

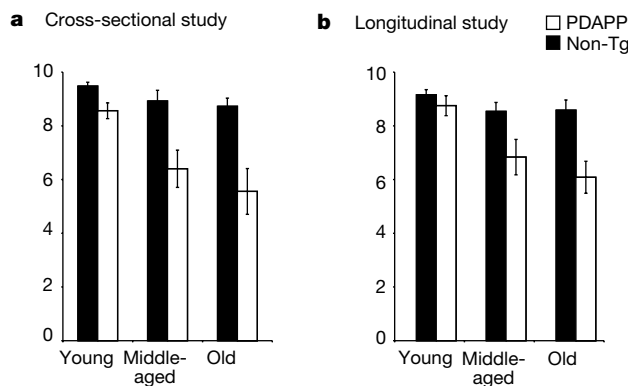


Figure 3 An age-related decline in learning capacity in PDAPP mice. The *y* axes show the number of successive spatial locations of a single hidden platform learned to criterion in 10 d of training ('learning capacity'). Older PDAPP mice performed significantly more poorly than older non-Tgs (interaction effect for cross-sectional: $F = 4.05$, d.f. 2/93, $P < 0.05$; longitudinal: $F = 6.83$, d.f. 2/46, $P < 0.005$). **a**, Cross-sectional study. There was a trend towards an age-related decline in learning capacity in non-Tgs ($F = 2.94$, d.f. 2/47, $0.10 > P > 0.05$), but the decline was highly significant in the PDAPP mice ($F = 8.84$, d.f. 2/46, $P < 0.005$). **b**, Longitudinal study. Learning capacity declines in the same age-related way in individual PDAPP mice retested throughout the lifespan. Non-Tgs again showed no change in learning capacity as a function of age ($F = 1.66$, d.f. 2/24, $P > 0.2$), whereas the PDAPP mice learned steadily fewer platform locations ($F = 18.30$, d.f. 2/22, $P < 0.001$). Results shown are mean $\pm$ 1 s.e.m.

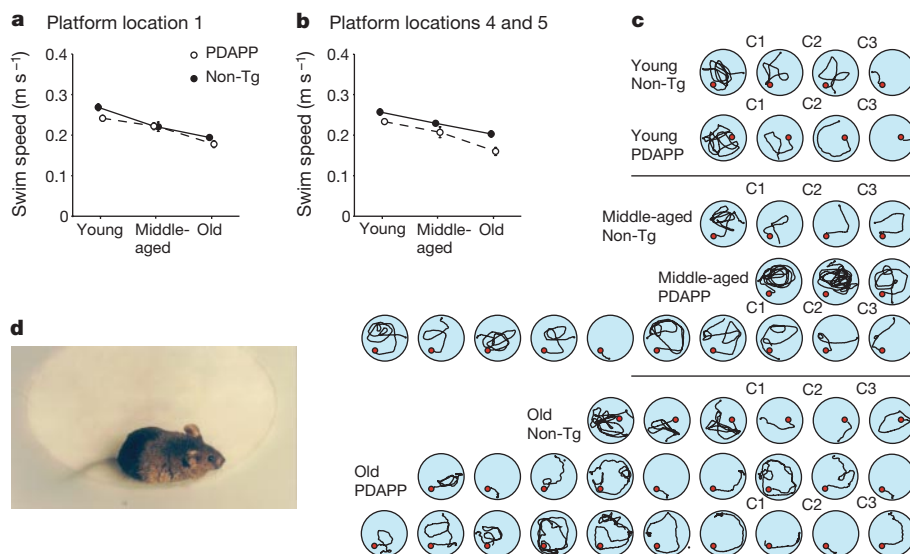


Figure 2 Swim speeds and swim paths. **a**, **b**, Mean ($\pm$ s.e.m.) swim speeds over the first three trials of training to the first, and the 4th and 5th platform locations. These trials were chosen because, with the performance criterion being three trials with an escape latency less than 20 s, the first three trials of training are only ones for which we have data from all animals. Swim-speed declines with age (location 1: $F = 27.46$, d.f. 2/93, $P < 0.001$; locations 4 and 5: $F = 26.67$, d.f. 2/93, $P < 0.001$). The small but significant difference between PDAPPs and non-Tgs (overall: 0.024 m s^{-1} ; location 1:

$0.10 > P > 0.05$; locations 4 and 5: $F = 18.91$, d.f. 1/93, $P < 0.001$) did not change as a function of age ($F < 1$). **c**, Paths taken by representative animals (that is, close to the group mean in terms of trials to criterion) in learning the 5th location at each of the three ages. Note similar patterns in PDAPP and non-Tg mice at 6–9 months, but circuitous paths taken by the PDAPP mice on some but not all trials at 13–15 and 18–21 months. **d**, A PDAPP mouse on the hidden escape platform in the water maze.

III entorhinal cells in the outer blade of the molecular layer of the dentate gyrus and stratum moleculare lacunosum of area CA1 (Fig. 4b, middle-aged animal). Occasional plaques were seen in cortex, particularly the cingulate cortex. By 21–22 months, amyloid plaques were apparent through many regions of the hippocampal formation, the cingulate cortex, insular cortex dorsal to the rhinal fissure and entorhinal cortex (Fig. 4c, old animal). A zone of cortex between these areas showed a lower density of amyloid plaques but a light background of diffuse staining. No amyloid plaques were observed in control mice. β -Amyloid plaque burden in the hippocampus increased as a function of age as described previously (Fig. 4d)^{8,14}. Behavioural performance declined as a function of plaque burden (Fig. 4e). A scatter plot of middle-aged and old mice only revealed a significant negative correlation between learning capacity and plaque burden (Fig. 4f).

In contrast to the learning of several spatial locations, object-recognition memory was normal in PDAPP animals across their lifespan. This was tested using a task requiring the spontaneous recognition of a series of new and familiar objects^{15,16}. We observed (cross-sectional study) that both PDAPP and non-Tg mice preferentially explored each of five separate new objects to the same extent (Fig. 5). The memory retention interval was varied to examine forgetting rates. Forgetting occurred at the same rate in PDAPP and non-Tg mice.

These findings indicate that the spatial learning impairment displayed by PDAPP mice can be dissociated into age-related and age-independent components. The age-related component became apparent through the use of a new task in which mice are required to learn a series of successive spatial locations. This task is sensitive to age in non-transgenic controls but, in PDAPP mice, the age-related

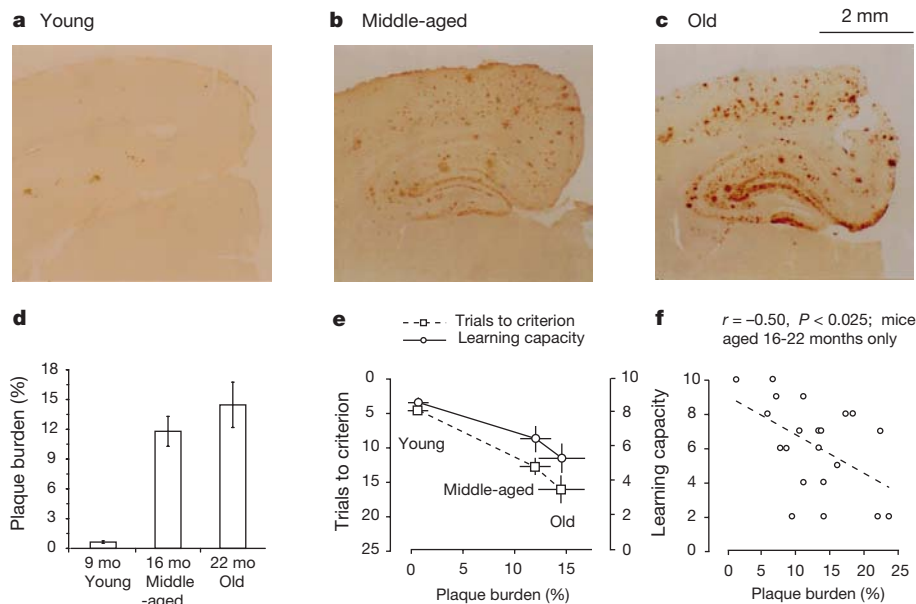


Figure 4 The relationship between performance and β -amyloid plaque deposition. **a–c**, Diffuse amyloid plaque deposition revealed using the 3D6 antibody in PDAPP mice of different ages. These sections are taken from the three mice whose paths are shown in Fig. 2. Note prominent deposition of plaques in the outer molecular layer of the dentate gyrus by age 16 months. **d**, Plot of plaque burden as a function of age measured in a subset of mice. Quantitative analysis of plaque burden in the hippocampus confirmed an age-related increase ($F = 13.4$, d.f. 2/24, $P < 0.001$). **e**, Two measures of performance in learning a series of spatial locations as a function of plaque burden. Left-hand y axis, trials to criterion on locations 4 and 5 (note axis is inverted); right-hand y axis, learning capacity (number of platform locations learned to criterion in 10 d). The behavioural scores are the means $\pm$ s.e.m. of all animals tested at each age. **f**, Scatter plot of learning

capacity scores in the subset of middle-aged and old animals for which plaque burden measures are available. An overall analysis of all 27 PDAPP mice revealed significant correlations between hippocampal plaque burden and both the number of trials to reach criterion for platform locations 4 and 5, and the separate measure of learning capacity ($r = 0.57$ and $r = -0.59$, respectively; $P < 0.01$). However, these correlations may be artificially inflated by the inclusion of young mice displaying few plaques. We also examined the middle-aged and old animals as a single group (age range 16–22 months; $n = 21$; these age groups did not differ significantly with respect to either plaque burden or behavioural performance, $P > 0.5$ in all cases, Bonferroni post-hoc comparisons). The correlation coefficients for these 21 animals were lower but were significant in both cases ($r = 0.43$ (trials to criterion) and $r = -0.50$ (learning capacity); $P < 0.05$).

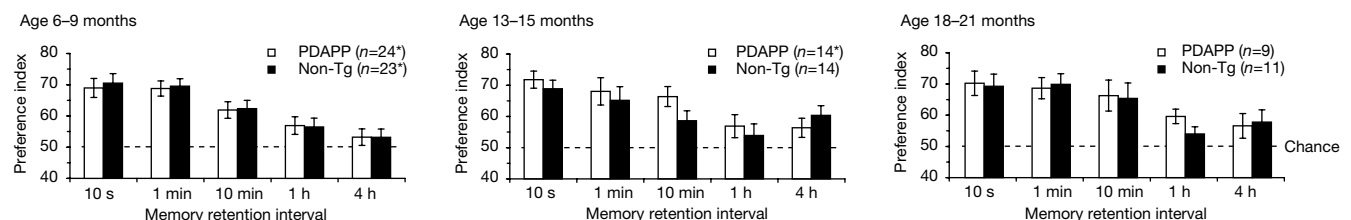


Figure 5 Normal object-recognition memory. Preference index for recognition of a familiar object as a function of memory delay interval, group and age (mean $\pm$ s.e.m.). Chance performance is 50% (dotted line). PDAPP and non-Tgs show equivalent above chance performance reflecting their greater exploration of the new object. The memory delay was varied across objects and familiarity observed to decay over time ($F = 20.8$, d.f. 4/352, $P < 0.001$) at the same rate in PDAPP and non-Tg mice ($F < 1$). The mean

performance of both groups was significantly above chance at all time intervals ($P < 0.05$ or better; one-sample t -tests). Comparable results were obtained in the longitudinal study (data not shown). Three young mice (1 PDAPP, 2 non-Tg) and one middle-aged PDAPP mouse failed to explore the objects in the groups marked with an asterisk and these were excluded from the data analysis. Results shown are mean $\pm$ s.e.m.

decline is substantially greater. The deficit in learning and recalling separate platform locations learned successively was apparent both when the PDAPP mice received their first training at the different ages tested, and when initially young PDAPP mice were re-tested through their lifespan. The use of a longitudinal design in the behavioural testing of hAPP mice is new, and reveals that individual PDAPP mice have, relative to age-matched controls, a declining capacity to learn and/or recall as they get older. This deficit in working- or episodic-like memory in the PDAPP mouse is consistent with the age-related deficit in T-maze alternation exhibited by APP₆₉₅SWE mice¹⁷.

The age-independent learning deficit in PDAPP mice was restricted to learning the first platform location, an impairment analogous to that seen in previous water-maze reference memory experiments²⁻⁷. In the single report claiming an age-related impairment in reference memory in APP₆₉₅SWE mice¹⁸, there is reasonable concern that an age-related change occurred in the control rather than the transgenic animals. This age-independent deficit is unrelated to progressive changes in β -amyloid levels and plaque deposition¹⁹, but may be associated with the smaller size of hippocampus and glucose hypometabolism that occurs in V717F mice at a young age^{20,21}.

A general implication is that dissociations between age-related and age-independent components of representational memory need to be taken into account in animal models of progressive neurodegenerative diseases if valid tests of the differential effects of A β overexpression, β -amyloid plaques and other neurodegenerative changes on cognitive function are to be made. We propose that as β -amyloid plaques are deposited, β -amyloid levels rise and fast synaptic transmission in the hippocampus and other brain areas becomes less effective^{19,22}, PDAPP mice are gradually less able to rapidly encode changes in their memory representation of an environment and/or to retrieve information selectively—a deficit in 'episodic-like' memory^{12,23}. These mice also show an early and age-independent deficit in spatial learning²⁴ but not in object-recognition memory (but see ref. 25).

The decline in performance in this new water-maze task correlates with the age-related increase in β -amyloid plaque deposition, even when young animals with minimal deposits are excluded. However, we do not yet know if this relationship to plaque deposition is causal for our findings are also consistent with the primary pathological trigger being elevated A β levels rather than frank plaque deposition²⁶. It will be valuable to use this new test to examine hAPP mice that have received exogenous treatments limiting the deposition of β -amyloid plaques²⁷. □

Methods

Animals

In the cross-sectional study, there were 99 PDAPP and non-transgenic littermate control mice (non-Tg) aged 6–9 months (young), 13–15 months (middle-aged) and 18–21 months (old). A subset of the young animals formed the longitudinal study, with repeat testing at the two later ages, and a total of 25 mice completing the study at age 22 months (PDAPP = 12, non-Tg = 13). All animals were male, of common parentage, on a Swiss Webster/DBA/C57BL6 background and fed freely on food and water.

Experimental protocol

Behavioural testing took place in a water maze and an exploratory arena. For cued navigation, testing began with 5 d of cued training to a visible platform. For training to criterion, place navigation to a single hidden platform was then tested, followed by training to four successive new locations. Each animal was trained, for up to eight trials per day, to a rigorous performance criterion of three successive trials with an escape latency of less than 20 s before being transferred to the next location on the next day. For object-recognition memory, testing took place in the arena over 5 d to examine rapidly acquired memory for object familiarity. For learning capacity, place navigation consisting of training to criterion of as many successive spatial locations as could be learned in a fixed 10-day training period. The only difference between the training-to-criterion and learning-capacity tests was that the former measured the extent of training required to learn five locations, and the latter measured the total number of platform locations that could be acquired over a set number of days. The animals were killed when all behavioural training was completed to measure the extent of β -amyloid plaque burden in the

hippocampus. The longitudinal component of the study involved only minor procedural modifications. On re-testing the mice when they became middle-aged and old, they were given 2 d of cued training, 3 d of object-recognition memory, and either 10 d or 5 successive platform locations (whichever took longer) of place-navigation training.

Water maze

Mice were trained in a 2-m diameter open-field water maze (opaque water at $25 \pm 1^\circ\text{C}$, 20-cm diameter platform, top surface 1.5 cm below water level, maximum trial duration = 90 s, 30 s on platform at end of trials). They were placed into the water at and facing the side walls, being transported there by hand from a holding cage in an adjacent room. Swim paths were monitored using an automated tracking system (Watermaze). The mice were run in squads of six at a time by two experimenters who were 'blind' with respect to their group assignment (three animals per squad were in each group). For cued training (four trials per day over 5 d), the pool was surrounded by white curtains to occlude sight of extra-maze cues. The platform was cued by means of a 15-cm high dark cylinder placed onto it and placed pseudo-randomly in different locations across trials. For the training-to-criterion and learning-capacity tests (eight trials per day), the curtains were pulled to the side to reveal extra-maze cues and the platform was unmarked by any local cues within the pool. For each mouse, the platform was moved between different locations, drawn from a set of 20 possible locations. The performance criterion for both tests was three trials in a row with an average escape latency of less than 20 s (maximum trials = 40 for location 1, and 32 trials thereafter).

Object recognition

The animals were tested in an arena with sloping side-walls to prevent shadows from overhead illumination (50 cm $\times$ 50 cm, 1 cm sawdust on floor). Habituation consisted of daily 5-min periods in the empty arena over 5 d. The mice were then allowed to explore two identical objects placed into the arena at fixed locations until they had accumulated 30 s of total inspection time (where this is defined as active exploration, sniffing, and so on, of the objects) or for a maximum of 20 min. Five different sets of identical objects were used on different days. After the initial period of inspection, the mice and the objects were removed from the arena for the prescribed memory delay interval (10 s, 1 min, 10 min, 1 h or 4 h spent in a transport cage in the laboratory). A third copy of the earlier identical objects was then placed into the arena at one location and a new object at the other. The mice were returned to the arena and allowed to explore until 30 s of inspection time had again accumulated. The toy objects (for example, a model duck) measured about 10–15 cm in height and were washed in alcohol between successive tests. Memory of the familiar object is associated with increased exploration of the new object and a performance index (PI) is calculated using the formula $\text{PI} = 100 \times (\text{new object inspection time})/30$.

Immunocytochemistry

The brains of a subset of mice ($n = 27$; young, $n = 6$; middle-aged, $n = 14$; old, $n = 7$) were removed and one hemisphere was drop-fixed in formalin. The tissue was mounted coronally and 30- μm sections collected using a cryostat. Every eighth section throughout the extent of the brain was stored in antifreeze solution (30% glycerol/30% ethylene glycol in 40 mM NaPO₄) at -20°C before immunostaining. The sections were incubated with 3D6-biotinylated antibody overnight at 4°C and then reacted with the horseradish peroxidase-avidin-biotin complex (Vector Laboratories) and developed using 3,3'-diaminobenzidine as the chromagen²⁸. Amyloid deposition was quantified automatically using a Leica Q-Win image analysis system. Video images of the brain were captured, the area of the hippocampus outlined and a threshold optical density was obtained that discriminated staining from background. The amyloid 'plaque burden' was defined as the average percentage of hippocampal area covered by amyloid depositions over six sections.

Received 5 October; accepted 22 November 2000.

- Morris, R. G. M., Garrud, P., Rawlins, J. N. P. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* **297**, 681–683 (1982).
- Yamaguchi, F. et al. Transgenic mice for the amyloid precursor protein 695 isoform have impaired spatial memory. *NeuroReport* **2**, 781–784 (1991).
- D'Hooge, R., Nagels, G., Westland, C. E., Muck, L. & De Deyn, P. P. Spatial learning deficit in mice expressing human 751-amino acid beta-amyloid precursor protein. *NeuroReport* **7**, 2807–2811 (1996).
- Nalbantoglu, J. et al. Impaired learning and LTP in mice expressing the carboxy terminus of the Alzheimer amyloid precursor protein. *Nature* **387**, 500–505 (1997).
- Justice, A. & Motter, R. Behavioral characterization of PDAPP transgenic Alzheimer mice. *Soc. Neurosci. Abstracts* **23**, 1637 (1997).
- Moechars, D. et al. Early phenotypic changes in transgenic mice that overexpress different mutants of amyloid precursor protein in brain. *J. Biol. Chem.* **274**, 6483–6492 (1999).
- Raber, J. et al. Apolipoprotein E and cognitive performance. *Nature* **404**, 352–354 (2000).
- Games, D. et al. Alzheimer-type neuropathology in transgenic mice overexpressing V717F β -amyloid precursor protein. *Nature* **373**, 523–527 (1995).
- Rawlins, J. N. P. Associations across time: The hippocampus as a temporary memory store. *Behav. Brain Sci.* **8**, 479–497 (1985).
- Whishaw, I. Q. & Tomie, J. A. Perseveration on place reversals in spatial swimming pool tasks: further evidence for place learning in hippocampal rats. *Hippocampus* **7**, 361–370 (1997).
- Steele, R. J. & Morris, R. G. M. Delay-dependent impairment of a matching to place task with chronic and intrahippocampal infusion of the NMDA antagonist D-AP5. *Hippocampus* **9**, 118–136 (1999).
- Wood, E. R., Dudchenko, P. A. & Eichenbaum, H. The global record of memory in hippocampal neuronal activity. *Nature* **397**, 613–616 (1999).
- Aggleton, J. P. & Brown, M. W. Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behav. Brain Sci.* **22**, 425–489 (1999).

14. Chen, K. S. *et al.* Neurodegenerative Alzheimer-like pathology in PDAPP 717V→F transgenic mice. *Prog. Brain Res.* **117**, 327–334 (1998).
15. Ennaceur, A. & Delacour, J. A new one-trial test for neurobiological studies of memory in rats. 1: Behavioral data. *Behav. Brain Res.* **31**, 47–59 (1988).
16. Tang, Y. P. *et al.* Genetic enhancement of learning and memory in mice. *Nature* **401**, 63–69 (1999).
17. Chapman, P. F. *et al.* Impaired synaptic plasticity and learning in aged amyloid precursor protein transgenic mice. *Nature Neurosci.* **2**, 271–276 (1999).
18. Hsiao, K. *et al.* Correlative memory deficits, A β elevation, and amyloid plaques in transgenic mice. *Science* **274**, 99–102 (1996).
19. Johnson-Wood, K. *et al.* Amyloid precursor protein processing and A β 42 deposition in a transgenic mouse model of Alzheimer disease. *Proc. Natl Acad. Sci. USA* **94**, 1550–1555 (1997).
20. Dodart, J.-C., Mathis, C., Bales, K. R., Paul, S. M. & Ungerer, A. Early regional cerebral glucose hypometabolism in transgenic mice overexpressing the V717F β -amyloid precursor protein. *Neurosci. Lett.* **277**, 49–52 (1999).
21. Dodart, J.-C. *et al.* Neuroanatomical abnormalities in behaviorally characterised APPV717F Transgenic Mice. *Neurobiol. Dis.* **7**, 71–85 (2000).
22. Larson, J., Lynch, G., Games, D. & Seubert, P. Alterations in synaptic transmission from young and aged PDAPP mice. *Brain Res.* **840**, 23–35 (1999).
23. Clayton, N. S. & Dickinson, A. What, where and when: Episodic-like memory during cache recovery by scrub jays. *Nature* **395**, 272–274 (1998).
24. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Clarendon, Oxford, 1978).
25. Dodart, J. C. *et al.* Behavioral disturbances in transgenic mice overexpressing the V717F β -amyloid precursor protein. *Behav. Neurosci.* **113**, 982–990 (1999).
26. Naslund, J. *et al.* Correlation between elevated levels of amyloid β -peptide in the brain and cognitive decline. *J. Am. Med. Assoc.* **283**, 1571–1577 (2000).
27. Schenk, D. *et al.* Immunization with amyloid- β attenuates Alzheimer-disease-like pathology in the PDAPP mouse. *Nature* **400**, 173–177 (1999).
28. Masliah, E. *et al.* Comparison of neurodegenerative pathology in transgenic mice overexpressing V717F β -amyloid precursor protein and Alzheimer's disease. *J. Neurosci.* **16**, 5795–5811 (1996).

Acknowledgements

This work was supported by grants from the MRC and the Cunningham Trust. We are indebted to Bill Nailon (CJD Surveillance Unit, Edinburgh) and Karen Khan (Elan Pharmaceuticals) for assistance with immunocytochemistry and image analysis.

Correspondence and requests for materials should be addressed to R.G.M.M. (e-mail: R.G.M.Morris@ed.ac.uk).

A β peptide immunization reduces behavioural impairment and plaques in a model of Alzheimer's disease

Christopher Janus*, Jacqueline Pearson*, JoAnne McLaurin*, Paul M. Mathews†, Ying Jiang†, Stephen D. Schmidt†, M. Azhar Chishti*, Patrick Horne*, Donna Heslin*, Janet French*, Howard T.J. Mount*, Ralph A. Nixon†, Marc Mercken‡, Catherine Bergeron§, Paul E. Fraser*, Peter St George-Hyslop*§ & David Westaway*

* Centre for Research in Neurodegenerative Diseases, Departments of Medicine, Laboratory Medicine and Pathobiology, and Medical Biophysics, University of Toronto, Tanz Neuroscience Building, 6 Queen's Park Crescent West, Toronto, Ontario M5S 3H2, Canada

† Nathan Kline Institute Center for Dementia Research, and New York University School of Medicine, 140 Old Orangeburg Road, Orangeburg, New York 10962, USA

‡ Janssen Research Foundation, Turnhoutseweg, 30, B-2340 Beerse, Belgium

§ Departments of Medicine (Division of Neurology) and Pathology, Toronto Western Hospital, University Health Network, Toronto, Ontario M5S 1A8, Canada

Much evidence indicates that abnormal processing and extracellular deposition of amyloid- β peptide (A β), a proteolytic derivative of the β -amyloid precursor protein (β APP), is central to the pathogenesis of Alzheimer's disease (reviewed in ref. 1). In the PDAPP transgenic mouse model of Alzheimer's disease, immunization with A β causes a marked reduction in burden of the brain amyloid^{2,3}. Evidence that A β immunization also reduces cognitive

dysfunction in murine models of Alzheimer's disease would support the hypothesis that abnormal A β processing is essential to the pathogenesis of Alzheimer's disease, and would encourage the development of other strategies directed at the 'amyloid cascade'. Here we show that A β immunization reduces both deposition of cerebral fibrillar A β and cognitive dysfunction in the TgCRND8 murine model of Alzheimer's disease without, however, altering total levels of A β in the brain. This implies that either a ~50% reduction in dense-cored A β plaques is sufficient to affect cognition, or that vaccination may modulate the activity/abundance of a small subpopulation of especially toxic A β species.

To explore the behavioural consequences of A β immunization, we used the TgCRND8 murine model of Alzheimer's disease that expresses a mutant (K670N/M671L and V717F) human β APP₆₉₅ transgene under the regulation of the Syrian hamster prion promoter on a C3H/B6 strain background (M.A.C. *et al.*, manuscript in preparation). TgCRND8 mice have spatial learning deficits at 3 months of age that are accompanied by both increasing levels of SDS-soluble A β and increasing numbers of A β -containing amyloid plaques in the brain. Age- and sex-matched TgCRND8 mice and non-Tg littermates in three cohorts were vaccinated at 6, 8, 12, 16 and 20 weeks with either A β ₄₂ or islet-associated polypeptide (IAPP), which has similar biophysical properties to A β but is associated with a non-central nervous system (CNS) amyloidosis. Both immunogens were in β -pleated-sheet conformation at the

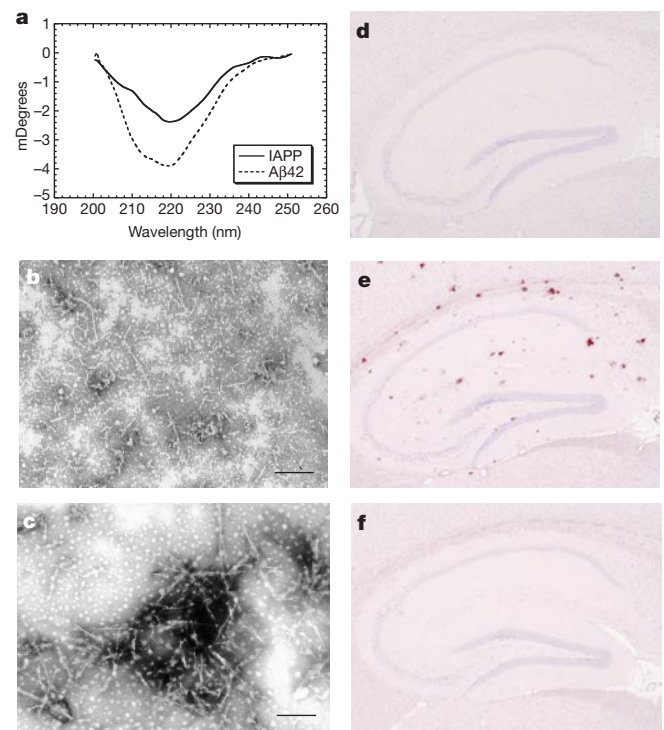


Figure 1 A β and IAPP peptide immunogens were predominantly β -structured and induced antibodies recognizing fibrillar A β deposits. **a**, Circular dichroism spectra of A β ₄₂ (dotted line) and IAPP (solid line) before immunization are predominantly β -structured. **mDegrees**, millidegrees. **b**, Negative-stain electron microscopy (scale 150 nm) of A β ₄₂ peptide immunogen showed varying length fibrils; **c**, IAPP peptide immunogen showed short laterally aggregated fibrils. **d, f**, Serum (1:1,000 dilution) from non-immunized TgCRND8 mice (**d**) and from 23-week old IAPP-immunized mice (**f**) did not recognize the A β plaques in adjacent, non-formic-acid treated sections from a non-immunized TgCRND8 mouse with abundant A β -positive plaques. **e**, Sera from A β ₄₂-immunized mice strongly labelled dense-cored amyloid plaques but not diffuse A β deposits (which are profusely present in these animals and were labelled with other anti-A β antibodies such as 4G8; ref. 28).

time of injection (Fig. 1) and induced detectable antibody titres in all mice by 13 weeks of age (as measured by enzyme-linked immunoabsorbent assay (ELISA) using fibrillar forms of the respective immunogen). These titres increased by a further ~2–3-fold at 23 weeks ($A\beta_{42}$ titres in $A\beta_{42}$ -immunized mice (mean $\pm$ s.e.m.): 1:3,640 $\pm$ 470 at 13 weeks; 1:7,500 $\pm$ 1,712 at 23 weeks; IAPP titres in IAPP-immunized mice: 1:3,833 $\pm$ 1,167 at 13 weeks; 1:11,500 $\pm$ 3661 at 23 weeks). The sera from $A\beta$ -immunized mice intensely decorated extracellular, dense-cored plaque deposits when applied as an immunohistochemical reagent to sections of brain from TgCRND8 mice containing abundant amyloid plaques (which predominantly display $A\beta$ in a β -sheet conformation) or to formic-acid-treated sections (which also display additional non- β -sheet $A\beta$ -epitopes) (Fig. 1). However, these sera reacted only very weakly with diffuse, non-fibrillar $A\beta$ deposits, which can be readily detected in these tissues by anti- $A\beta$ monoclonal antibodies such as

4G8 (data not shown). The $A\beta$ -immune sera did not stain normal neurons, indicating limited crossreactivity with β APP holoprotein. In contrast, sera from non-immunized or IAPP-immunized mice did not stain any structures. Together, these data indicate that in this strain of mice immunization with $A\beta_{42}$ protofibrillar assemblies induced antibodies directed primarily towards $A\beta$ in a β -sheet conformation.

The mice were tested longitudinally in a reference memory version of the Morris water maze test at 11, 15, 19 and 23 weeks (Fig. 2). At each age of testing, the hidden platform was placed in a different quadrant of the pool. These data were analysed for the entire test period using a mixed model analysis of variance (ANOVA), with immunogen ($A\beta_{42}$ versus IAPP) and genotype (TgCRND8 versus non-Tg) as a between-subject factor, and age-of-testing (11, 15, 19 and 23 weeks) as a within-subject factor. This longitudinal design and mode of analysis, which simulates longitudinal human clinical trials⁴, revealed that $A\beta_{42}$ -immunized TgCRND8 mice performed significantly better than IAPP-immunized TgCRND8 mice ($P < 0.05$), with 31% of the performance variance being due to the effects of the immunogen. However, the improvement was partial—the $A\beta_{42}$ -immunized TgCRND8 mice did not perform as well as their non-Tg littermates ($P < 0.01$). Because the experimental design involved testing of naive mice at 11 weeks of age, followed by a series of reversal tests at 15, 19 and 23 weeks, an additional analysis carried out on the reversal tests confirmed improved performance for the $A\beta$ -immunized Tg mice ($P < 0.02$) during this phase of testing as well. The overall conclusion that $A\beta_{42}$ immunization ameliorates the cognitive deficit of TgCRND8 mice was robust, regardless of whether the analysis assessed latency to reach the hidden platform or swim path length (a measure that is less sensitive to swim speed and floating⁵).

The improved performance of $A\beta_{42}$ -immunized TgCRND8 mice was not due to a nonspecific effect of immunization or to an effect on other behavioural, motor, or perceptual systems. Control studies

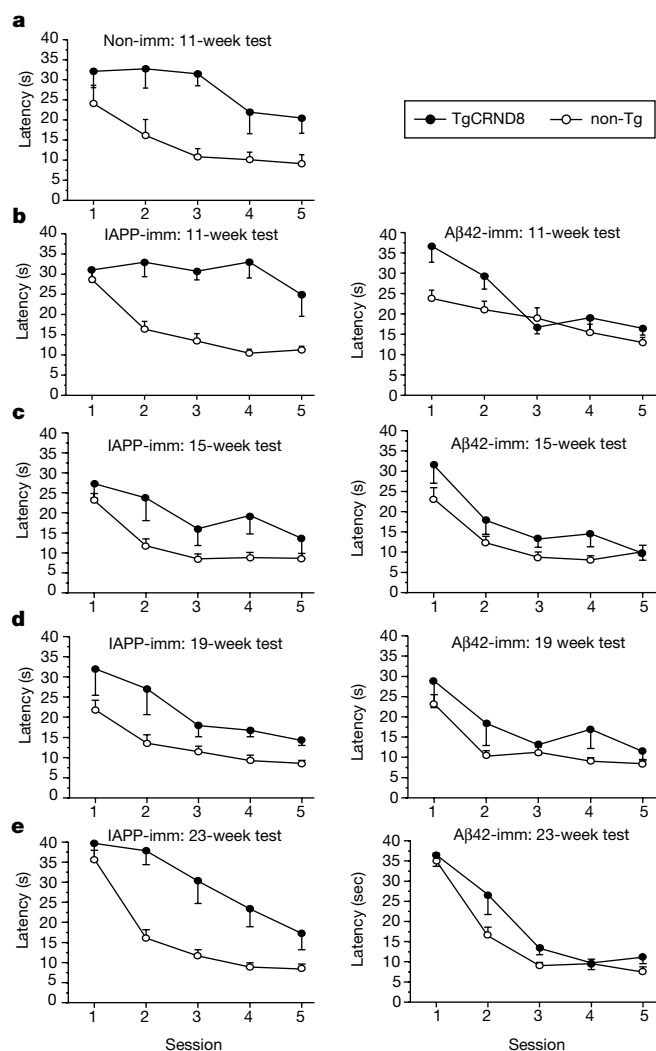


Figure 2 Reference memory version of Morris water maze test in TgCRND8 mice. At 11 weeks of age, non-immunized TgCRND8 mice ($n = 5$) show cognitive impairment relative to non-Tg controls ($n = 8$) (a), which is similar to that of IAPP-immunized TgCRND8 mice ($n = 12$) (b, left), whereas the performance of $A\beta_{42}$ -immunized TgCRND8 mice ($n = 9$) (b, right) approaches that of non-Tg littermates ($n = 19$). At 15 (c) and 19 (d) weeks of age, the IAPP-immunized TgCRND8 mice ($n = 6$) (left) were impaired compared with non-Tg littermates ($P < 0.01$, $\omega^2 \approx 36\%$; $n \approx 15$), but were not significantly different from the $A\beta_{42}$ -immunized TgCRND8 mice ($n \approx 6$) (right). At 23 weeks of age, the IAPP-immunized TgCRND8 mice ($n = 6$) (e, left) were significantly impaired relative to both non-Tg littermates ($n = 15$; $P < 0.001$, $\omega^2 = 65\%$) and $A\beta_{42}$ -immunized TgCRND8 mice ($n = 6$; $P < 0.01$) (e, right). Vertical bars represent s.e.m. See text and Supplementary Information for statistical analyses.

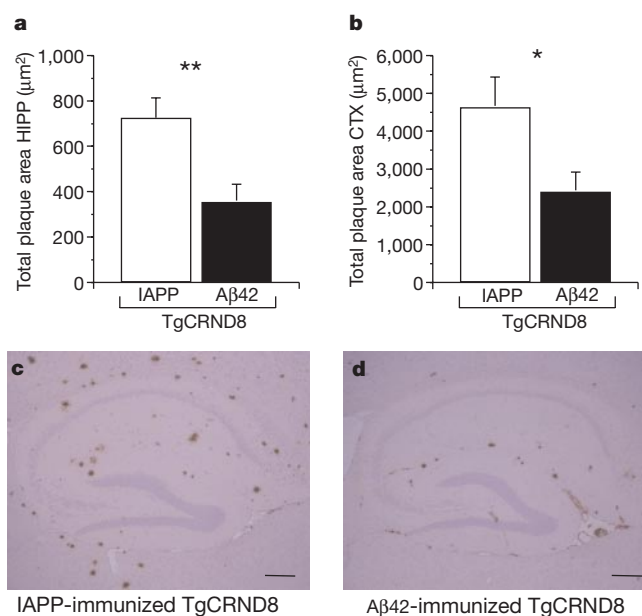


Figure 3 At 25 weeks of age, in TgCRND8 mice immunized with $A\beta_{42}$ peptides, dense-cored $A\beta$ plaque burden is reduced in the hippocampus (HIPP) (a) and in the cerebral cortex (CTX) (b). The $A\beta_{42}$ -immunized TgCRND8 mice had 50% fewer plaques than IAPP-immunized TgCRND8 controls (71.4 ± 10.8 per area counted versus 119.7 ± 14.6 in the cortex, $P < 0.05$; and 11.6 ± 1.6 per area counted versus 20.9 ± 1.7 in hippocampus, $P < 0.01$). Representative pictures of the distribution of $A\beta$ plaques labelled by the Dako 6F/3D anti- $A\beta$ monoclonal antibody in hippocampus of IAPP- and $A\beta_{42}$ -immunized TgCRND8 mice (c and d, respectively). Vertical bars represent s.e.m. Asterisk, $P < 0.05$; two asterisks, $P < 0.01$. Scale bars, 100 μ m.

in non-Tg mice established that immunization with Freund's Adjuvant plus phosphate buffered saline, Freund's plus A β_{42} , Freund's plus IAPP, IAPP alone, or A β_{42} alone had no effect on performance in the Morris water maze test ($P > 0.1$). During non-spatial pre-training at 11 weeks, the latency of random search for a hidden platform, as well as swim speed, and swim path length to a visible platform were not affected by genotype or immunization ($P > 0.05$). Similarly, at 23 weeks, performance during a visible platform test and spontaneous exploration of the open field were not significantly influenced by immunization ($P > 0.05$). Gender effects were not significant between the TgCRND8 and IAPP- or A β_{42} -immunized non-Tg littermates ($P > 0.05$).

Analysis of performance in the probe trials with the platform removed (annulus crossing index, passes over platform site or dwelling in the target quadrant) revealed no significant differences owing to immunization or genotype, or interactions between these two factors (each $P > 0.05$). This probably arises both because the probe trials were conducted within 30 min of the final training trial on day 5 (by which point most mice had learned), and because repeated administration of probe trials (at 15, 19 and 23 weeks) makes them a less effective measure of spatial memory⁶.

Because the main analysis revealed significant immunogen $\times$ genotype ($P < 0.01$) and immunogen $\times$ genotype $\times$ age ($P < 0.05$) interactions, as well as significant main factor effects for immunogen, genotype and age of testing ($P < 0.01$ for all), *post hoc* analyses were carried out for each age of testing. These showed markedly improved performance in A β_{42} -immunized TgCRND8 mice at 11 and 23 weeks, with A β_{42} -immunization accounting for large portions of the variance (ω^2) (19% at 11 weeks; 42% at 23 weeks). Analyses at 15 and 19 weeks did not show statistically significant differentiation of the two immunogens. However, this probably reflects the effects of previous test experience ('carrying-over' effect⁷). Reduction of impairment during re-testing in the water maze has been reported in another study of a murine model of Alzheimer's disease⁸, and similar re-test effects are commonly seen in placebo-treated patients in human trials⁹. The subsequent re-emergence of impaired cognition in the IAPP-immunized TgCRND8 mice at 23 weeks probably reflects the advance of Alzheimer's-disease-related phenotypes, which are clearly progressive over this epoch in TgCRND8 mice. Thus, increased A β levels are first detectable by western blot or ELISA analysis in brain homogenates at 8.5–17 and 10 weeks, respectively, but thereafter rise progressively. Similarly, cortical dense-cored A β -plaques are first detectable immunohistochemically at 7–10 weeks, but increase by a further 18-fold at 19–27 weeks (M.A.C. *et al.*, manuscript in preparation).

In agreement with previous reports^{2,3}, A β_{42} immunization caused ~50% reduction in the number and size of A β -positive dense-cored plaques (Fig. 3) without affecting steady-state levels of β APP holoprotein, amino-terminal secreted fragments (β APP_s), or carboxy-terminal fragments in the brain (see Supplementary Information). However, in contrast to the previous reports, A β_{42} immunization had no significant effect on the levels of formic-acid-extractable A β in brain at 13 weeks ($P > 0.1$) or 25 weeks (IAPP immunized ($n = 7$): A $\beta_{40} = 11,475 \pm 1,567$ fmol per mg protein, A $\beta_{42} = 39,983 \pm 4,387$; A β_{42} immunized ($n = 6$): A $\beta_{40} = 12,276 \pm 2,386$; and A $\beta_{42} = 50,457 \pm 8,552$, mean $\pm$ s.e.m., $P \approx 0.27$ two-tailed *t*-test). Sandwich ELISAs using two other detection antibodies (directed at A β residues 1–4 and 17–24) yielded similar results, arguing against a selective effect of immunization on A β species differing in their amino terminus identity or in post-translational modification.

One explanation for our results is that dense-cored cerebral amyloid plaques are the toxic moiety, and the ~50% reduction caused by A β_{42} vaccination is sufficient to prevent or reverse the behavioural deficits. Although this cannot be excluded, the density of such plaques correlates poorly with the ante-mortem severity of dementia in most human studies¹⁰. Furthermore, a dissociation

between plaque deposition and cognitive and/or neuronal dysfunction has been described in other TgAPP mice^{11,12}. An alternative explanation is that immunization affects A β either in a particular conformation (for example, β -sheet forms in protofibrils) or in a restricted compartment. The former is more likely because we used oligomeric assemblies of A β in β -sheets ('protofibrils') as an immunogen, and the resultant antisera preferentially recognized β -sheet forms of A β . This is significant because monoclonal antibodies raised to A β epitopes that initiate fibril aggregation inhibit assembly of synthetic A β oligomeric protofibrils *in vitro*¹³. It is possible, therefore, that the antibodies induced in the current strain of mice may bind to β -sheet oligomeric aggregates and inhibit further assembly. This A β species is especially neurotoxic, a critical intermediary in fibrillogenesis and an accurate predictor of neurodegeneration^{14–17}. Consequently, small amounts of such antibodies that cross the blood–brain barrier (0.1% of serum levels³) might be sufficient to attenuate both the behavioural deficits caused by this neurotoxic form of A β and the further aggregation of these species into fibrillar A β in dense-cored plaques. Because this pool of A β is small, and because antibodies to this form of A β might need only inhibit assembly of A β fibrils to have a functional effect, these antibodies need not necessarily cause large changes in total cerebral A β . It is also possible that the antibodies redistribute A β from dense-cored plaques to diffuse A β deposits. Either might explain the divergent effects on levels of dense-cored plaques and A β measured biochemically (although high levels of cerebral A β in the presence of relatively few plaques are seen in humans¹⁸ and in transgenic mouse models¹⁹).

It is conceivable, however, that immunization might modulate A β metabolism through several distinct mechanisms, including destruction of A β by microglial phagocytosis³. Such different effects, which might reflect variations in antigen presentation or in strain-specific immune response, might also explain the disparity between the effects of A β immunization on total A β levels in this and the previous studies. If correct, such variations could complicate the use of active immunization in humans. In addition to clinical caveats raised previously²⁰, it is also important to emphasize that, although A β_{42} immunization had a strong effect on behaviour and neuropathology in this mouse model, it did not fully reverse these features. This might reflect either inefficient ingress of antibodies to the CNS and/or the possibility that other β APP proteolytic fragments may be involved in the pathogenesis of Alzheimer's disease²¹. These issues will need to be addressed directly by future studies. Nonetheless, our data support the hypotheses that A β plays a central role in Alzheimer's disease and that procedures that modulate its production, assembly and/or removal might be used as treatments. $\square$

Methods

Mice

The TgCRND8 mice (M.A.C. *et al.*, manuscript in preparation) were maintained in a hybrid C3H/B6 background. Experimental groups derived from crosses to B6 mice were matched for gender and weight, and transgene identity was unknown to experimenters at all stages of the study. Sixty-eight 6-week-old mice were immunized (TgCRND8, $n = 28$, non-Tg, $n = 40$); 60 entered the behavioural testing at 11 weeks (A β immunized: TgCRND8, $n = 12$; non-Tg, $n = 20$; and IAPP immunized: TgCRND8, $n = 9$; non-Tg, $n = 19$). Thirteen mice were killed at 13 weeks, and their brains analysed (A β immunized: TgCRND8, $n = 3$; non-Tg, $n = 4$; IAPP immunized: TgCRND8, $n = 2$, non-Tg = 4). All remaining mice were longitudinally tested until 23 weeks of age (A β immunized: TgCRND8, $n = 9$; non-Tg, $n = 16$; IAPP immunized: TgCRND8, $n = 7$; non-Tg, $n = 15$). Thirteen non-immunized mice (TgCRND8, $n = 5$; non-Tg, $n = 8$) were included as controls in the test conducted at 11 weeks of age.

Immunization procedures

Synthetic A β_{42} and IAPP peptides were isolated by reverse phase high performance liquid chromatography on a C₁₈ μ bondapak column, with purity determined using mass spectrometry and amino-acid analyses. Before preparation of the vaccine, the secondary structure and fibre morphology of the peptides were assessed using circular dichroism spectroscopy and by electron microscopy²². The immunization protocol and schedule have been described². Antibody titres were assayed in triplicate by ELISA (see

Supplementary Information for more details) in serum samples (200 µl of blood) collected at 13 and 25 weeks.

Behavioural tests and data analysis

The water maze apparatus, mouse handling and general testing procedures have been described²³. Before the first spatial learning test at 11 weeks, all mice underwent non-spatial pre-training (NSP) to assess swimming abilities and to accustom mice to the test^{24,25} (see Supplementary Information). Two days after the NSP phase, all mice underwent a reference memory training with a hidden platform placed in the centre of one quadrant of the pool for 5 days, with four trials per day. After the last trial of day 5, the platform was removed from the pool and each mouse received one 60-s swim probe trial. Escape latency (s), length of swim path (cm), swim speed (cm s⁻¹), % of floating (speed less than 5 cm s⁻¹), % of time in outer zone (near the pool wall), and % of time and path in each quadrant of the pool were recorded using an on-line HVS image video tracking system²³ (see Supplementary Information).

For the probe trials, an annulus-crossing index was calculated that represents the number of passes over the platform site, minus the mean of passes over alternative sites in other quadrants. The index expresses the spatial place preference and controls for alternative search strategies without place preferences, such as circular search paths^{26,27}. All mice were re-tested at 15, 19 and 23 weeks of age, one week before the next immunization. At each re-testing, the platform was placed in the centre of a different, semi-randomly chosen pool quadrant for all five sessions of training. At the end of the experiment, all mice were given a cue (visual platform) learning test. This was followed by the open-field test to investigate spontaneous locomotor exploration. Behavioural data was analysed using a mixed model of factorial ANOVA. Degrees of freedom were adjusted by Greenhouse–Geisser epsilon correction for heterogeneity of variance. A Bonferroni Inequality correction was applied for multiple comparisons. Omega squared (ω^2) was used as a measure of effect size caused by different factors.

Analysis of β APP and amyloid burden in brain

Three 5-µm sections at 25-µm intervals from one cerebral hemisphere were immunostained with Dako 6F/3D anti-A β monoclonal antibody to residues 8–17 (which is primarily reactive against dense-cored plaques) with 4G8 (ref. 28), or with sera from immunized mice, and counterstained with haematoxylin and resin mounted as described (M.A.C. *et al.*, manuscript in preparation). For some samples the formic-acid treatment step was omitted. End products were visualized with diaminobenzidine. Amyloid plaque burden was assessed using Leco IA-3001 image analysis software interfaced with a Leica microscope and a Hitachi KP-M1U CCD video camera. The quantitative analysis was performed at $\times 25$ magnification, and the image frame and guard size was set to 0,0,639,479 (307,200 µm²) for each slide. The brain area (cortex or hippocampus) was outlined using the edit plane function, and the area and number of plaques in the outlined structure were recorded. Data were pooled for all three sections.

Cerebral A β levels were assayed from formic-acid-extracted²⁹, hemi-brain sucrose homogenates using an ELISA method (see Supplementary Information) in which A β was trapped with either monoclonal antibody to A β ₄₀ (JRF/cAb40/10) or A β ₄₂ (JRF/cAb42/26) and then detected with horseradish peroxidase (HRP)-conjugated JRF/Ab40/17. The dilution of JRF/Ab40/17 and samples were optimized to detect A β in the range of 50 to 800 fmol ml⁻¹. ELISA signals are reported as the mean $\pm$ s.e.m. of four replica wells in fmol A β per mg total protein (determined with the BioRad DC protein assay), based on standard curves using synthetic A β _{1–40} and A β _{1–42} peptide standards (American Peptide Co. Sunnyvale, CA). Cerebral β APPs levels were analysed in supernatant of brain as described³⁰.

Received 19 July; accepted 16 November 2000.

- Steiner, H., Capell, A., Leimer, U. & Haass, C. Genes and mechanisms involved in beta-amyloid generation and Alzheimer's disease. *Eur. Arch. Psychiatr. Clin. Neurosci.* **249**, 266–270 (1999).
- Schenk, D. *et al.* Immunization with A β attenuates Alzheimer's disease-like pathology in the PDAPP mouse. *Nature* **400**, 173–177 (1999).
- Bard, F. *et al.* Peripherally administered antibodies against amyloid beta-peptide enter the central nervous system and reduce pathology in a mouse model of Alzheimer disease. *Nature Med.* **6**, 916–919 (2000).
- Whitehouse, P. J. *et al.* Clinical trial designs for demonstrating disease-course-altering effects in dementia. *Alzheimer Dis. Assoc. Disord.* **12**, 281–294 (1998).
- Morris, R. Developments of a water-maze procedure for studying spatial learning in the rat. *J. Neurosci. Methods* **11**, 47–60 (1984).
- Markowska, A. L., Long, J. M., Johnson, C. T. & Olton, D. S. Variable-interval probe test as a tool for repeated measurements of spatial memory in the water maze. *Behav. Neurosci.* **107**, 627–632 (1993).
- Martin, P. & Bateson, P. *Measuring Behaviour* (Cambridge Univ. Press, Cambridge, 1996).
- Hsiao, K. Correlative memory deficits, A β elevation, and amyloid plaques in transgenic mice. *Science* **274**, 99–102 (1996).
- Rogers, S. L., Farlow, M. R., Doody, R. S., Mohs, R. & Friedhoff, L. T. A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. Donepezil study group. *Neurology* **50**, 136–145 (1998).
- Terry, R. D., Masliah, E. & Hansen, L. A. in *Alzheimer Disease* 2nd edn (eds Terry, R. D., Katzman, R., Bick, K. C. & Sisodia, S. S.) (Lippincott Williams Wilkins, Philadelphia, 1999).
- Moechars, D. *et al.* Early phenotypic changes in transgenic mice that over-express different mutants of amyloid precursor protein. *J. Biol. Chem.* **274**, 6483–6492 (1999).
- Hsia, A. Y. Plaque-independent disruption of neural circuits in Alzheimer's disease mouse models. *Proc. Natl Acad. Sci. USA* **96**, 3228–3233 (1999).
- Solomon, B., Koppel, R., Hanan, E. & Katzav, T. Monoclonal antibodies inhibit in vitro fibrillar aggregation of the Alzheimer β -amyloid peptide. *Proc. Natl Acad. Sci. USA* **93**, 452–455 (1996).
- Lorenzo, A. & Yanker, B. A. β -amyloid neurotoxicity requires fibril formation and is inhibited by

Congo red. *Proc. Natl Acad. Sci. USA* **91**, 12243–12247 (1994).

- Walsh, D. M. *et al.* Amyloid β -protein fibrillogenesis. Structure and biological activity of protofibrillar intermediates. *J. Biol. Chem.* **274**, 25945–25952 (1999).
- Hartley, D. M. Protofibrillar intermediates of amyloid β -protein induce acute electrophysiological changes and progressive neurotoxicity in cortical neurons. *J. Neurosci.* **19**, 8876–8884 (1999).
- McLean, C. A. *et al.* Soluble pool of A β amyloid as a determinant of severity of neurodegeneration in Alzheimer's disease. *Ann. Neurol.* **46**, 860–866 (1999).
- Naslund, J. *et al.* Correlation between elevated levels of amyloid β -peptide in the brain and cognitive decline. *J. Am. Med. Assoc.* **283**, 1571–1577 (2000).
- Mucke, L. *et al.* High-level neuronal expression of A β 1–42 in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation. *J. Neurosci.* **20**, 4050–4058 (2000).
- St George-Hyslop, P. & Westaway, D. Alzheimer's disease: antibody clears senile plaques. *Nature* **400**, 116–117 (1999).
- Lu, D. C. A second cytotoxic proteolytic peptide derived from amyloid β -protein precursor. *Nature Med.* **6**, 397–404 (2000).
- McLaurin, J., Franklin, T., Chakrabarty, A. & Fraser, P. E. Phosphatidylinositol and inositol involvement in Alzheimer amyloid-beta fibril growth and arrest. *J. Mol. Biol.* **278**, 183–194 (1998).
- Janus, C. Spatial learning in transgenic mice expressing human presenilin 1 (PS1) transgenes. *Neurobiology of Aging* **21**, 541–549 (2000).
- Cain, D. P., Beiko, J. & Boon, F. Navigation in the water maze: the role of proximal and visual cues, path integration, and magnetic field information. *Psychobiology* **25**, 286–293 (1997).
- Dudchenko, P. A., Goodridge, J. P., Seiterle, D. A. & Taube, J. S. Effects of repeated disorientation on the acquisition of spatial tasks in rats: dissociation between the appetitive radial arm maze and aversive water maze. *J. Exp. Psychol.* **23**, 194–210 (1997).
- Gass, P. *et al.* Deficits in memory tasks of mice with CREB mutations depend on gene dosage. *Learn. Mem.* **5**, 274–288 (1998).
- Wehner, J. M., Sleight, S. & Upchurch, M. Hippocampal protein kinase C activity is reduced in poor spatial learners. *Brain Res.* **523**, 181–187 (1990).
- Wisniewski, H. M., Wen, G. Y. & Kim, K. S. Comparison of four staining methods on the detection of neuritic plaques. *Acta Neuropathol.* **78**, 22–27 (1989).
- Susuki, N. *et al.* An increased percentage of long amyloid B protein secreted by Familial Amyloid β Protein Precursor (BAPP717) mutants. *Science* **264**, 1336–1340 (1994).
- Citron, M. *et al.* Mutant presenilins of Alzheimer's Disease increase production of 42 residue amyloid β -protein in both transfected cells and transgenic mice. *Nature Med.* **3**, 67–72 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by the Medical Research Council of Canada, Howard Hughes Medical Research Foundation, Alzheimer Society of Ontario, The W. Garfield Weston Foundation and the US National Institute of Aging. We thank G. Carlson for useful discussions, and R. Renlund, K. Parisien, J. Haight and J. Cowieson for help during mice immunization. None of the authors has a financial or other relationship with Elan Pharmaceuticals Inc.

Correspondence and requests for materials should be addressed to P.H.StG.-H. (e-mail: p.hyslop@utoronto.ca).

A β peptide vaccination prevents memory loss in an animal model of Alzheimer's disease

Dave Morgan*, **David M. Diamond†‡**, **Paul E. Gottschall***, **Kenneth E. Ugen§**, **Chad Dickey§**, **John Hardy||**, **Karen Duff‡**, **Paul Jantzen***, **Giovanni DiCarlo***, **Donna Wilcock***, **Karen Connor***, **Jaime Hatcher#**, **Caroline Hope#**, **Marcia Gordon*** & **Gary W. Arendash#**

* Alzheimer Research Laboratory, Department of Pharmacology, † Department of Psychology, § Department of Medical Microbiology and Immunology, # Alzheimer Research Laboratory, Department of Biology, University of South Florida, Tampa, Florida 33612, USA

‡ James A. Haley VA Center, Tampa, Florida 33612, USA

|| Department of Pharmacology, Mayo Clinic, Jacksonville, Florida 32224, USA

§ Nathan Kline Institute, Orangeburg, New York 10962, USA

Vaccinations with amyloid- β peptide (A β) can dramatically reduce amyloid deposition in a transgenic mouse model of Alzheimer's disease¹. To determine if the vaccinations had deleterious or beneficial functional consequences, we tested eight months of A β vaccination in a different transgenic model for

Alzheimer's disease in which mice develop learning deficits as amyloid accumulates^{2,3}. Here we show that vaccination with A β protects transgenic mice from the learning and age-related memory deficits that normally occur in this mouse model for Alzheimer's disease. During testing for potential deleterious effects of the vaccine, all mice performed superbly on the radial-arm water-maze test of working memory. Later, at an age when untreated transgenic mice show memory deficits, the A β -vaccinated transgenic mice showed cognitive performance superior to that of the control transgenic mice and, ultimately, performed as well as nontransgenic mice. The A β -vaccinated mice also had a partial reduction in amyloid burden at the end of the study. This therapeutic approach may thus prevent and, possibly, treat Alzheimer's dementia.

The accumulation of fibrils formed from the A β peptide into

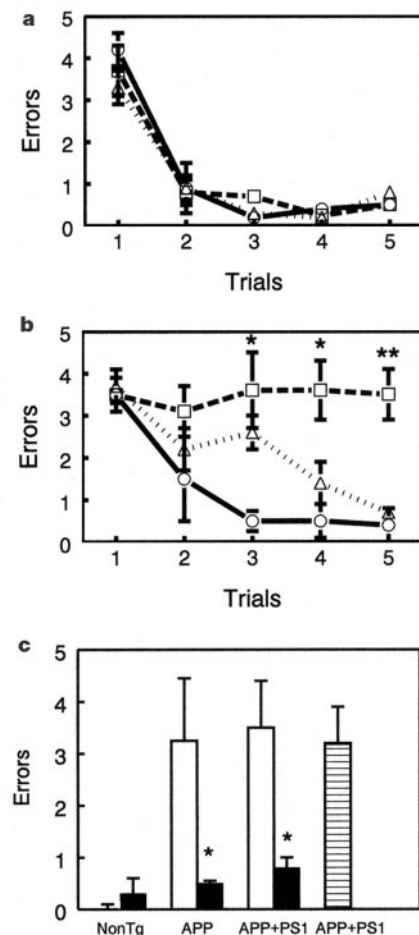


Figure 1 Radial-arm water-maze performance in vaccinated transgenic and nontransgenic mice. **a**, Nontransgenic mice (circles, solid lines), transgenic mice vaccinated with KLH (squares, dashed lines), and transgenic mice vaccinated with A β (triangles, dotted lines) were tested in the radial-arm water maze at 11.5 months of age (after five inoculations). All groups learned (trial 4) and remembered (trial 5) the platform location at this time point. In the same mice at 15.5 months of age (nine inoculations; **b**), the transgenic mice vaccinated with A β continued to show learning and memory of the platform location, whereas the transgenic mice vaccinated with KLH failed to show learning and memory for platform location on either trials 4 or 5 (* $P < 0.05$, ** $P < 0.01$; KLH significantly different from other two groups by LSD post hoc analysis after MANOVA). This benefit of A β vaccination was found in both the APP-only and APP+PS1 transgenic mice (**c**), with significantly fewer errors on trial 5 in the A β -vaccinated groups (solid bars) than in the KLH-vaccinated group (open bars) of both genotypes (* $P < 0.03$). Included for comparison is the trial 5 performance of another group (hatched bars) of untreated 15–16-month-old transgenic mice that were tested separately, and are reported on fully elsewhere².

amyloid plaques is a defining characteristic of Alzheimer's disease (AD). The A β vaccination protocol described in ref. 1 reduced A β deposits, which suggested that this approach might benefit AD patients. However, the functional consequences of such vaccinations might be deleterious. For example, plaque-associated inflammation promoted by the immunization could interfere with normal brain functioning, and/or lead to degenerative changes in the brain^{4–8}. We used a novel working-memory task that combines elements of a radial-arm maze and a water maze. This radial-arm water maze is remarkably robust at detecting learning/memory deficits that develop in AD transgenic mice² and more efficient in sample size requirements than other memory tasks typically used for rodents³.

To test the possibility that vaccinations might cause premature memory deficits in AD transgenic mice, we assessed learning/memory performance in the mice at 11.5 months of age after five inoculations with A β or the control vaccine, keyhole limpet haemocyanin (KLH). All mice showed strong learning and memory capacity, irrespective of treatment or transgene status (Fig. 1a). All groups averaged three to four errors on the first trial as they sought out the new platform location for that day, but averaged less than one error by trials 4 or 5, demonstrating intact working memory for platform location between trials and during the 30-min delay before trial 5. This strong performance by A β -vaccinated mice indicates that any inflammatory responses caused by the vaccine were not deleterious behaviourally.

Monthly inoculations were continued until the mice were 15.5 months, when these mice were tested again in the radial-arm water maze. At 15.5 months the KLH-vaccinated transgenic mice failed to demonstrate learning or memory of the platform location; their performance on all trials was the same (Fig. 1b). This is identical to the performance of other untreated transgenic mice that had been previously tested in this learning task at this age (Fig. 1c; ref. 3). In contrast, the A β -inoculated transgenic mice, although slower to learn platform location than nontransgenic mice on trial 3 (Fischer's least

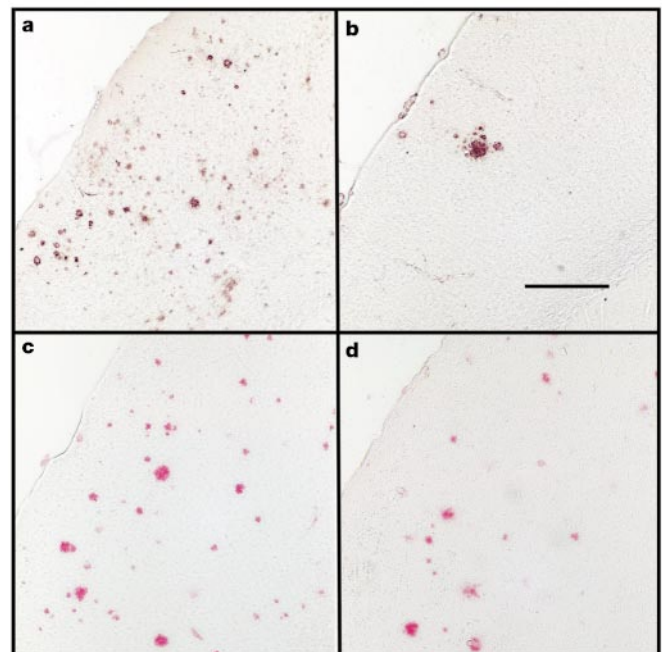


Figure 2 Amyloid pathology in transgenic mice vaccinated with KLH or A β . Immunohistochemistry for A β in frontal cortex is shown in (KLH-vaccinated) (**a**) and (A β -vaccinated) (**b**) in transgenic mice with values similar to the means shown in Fig. 3c. Congo-red staining is shown in (KLH) (**c**) and (A β) (**d**) in mice with values corresponding to the means in Fig. 3b. Horizontal sections are oriented with the corpus callosum in the lower right corner and anterior to the top. Scale bar, 500 μ m.

significant difference (LSD), $P < 0.02$), were nearly flawless by trial 5, and performed significantly better than the KLH-vaccinated transgenic mice on both trials 4 and 5 (multiple analysis of variance, MANOVA: $F_{(2,15)} = 5.83$, $P < 0.02$ and $F_{(2,15)} = 12.16$, $P < 0.001$, respectively; KLH transgenic group different from both other groups by Fischer's LSD post hoc comparisons, $P < 0.05$ on trial 4 and $P < 0.01$ on trial 5). Our individual evaluation of the performance of the two transgenic genotypes made it clear that both APP-only and APP+PS1 transgenic mice benefited from the A β vaccinations (Fig. 1c).

Serological analysis indicated that mice injected with A β developed antibodies against the A β peptide. Very high titres were found in both transgenic and nontransgenic mice immunized with A β ($IC_{50} = 27,000 \pm 5,000$ and $48,000 \pm 18,000$, respectively; not significant). There was no anti-A β activity in the KLH-immunized transgenic mice, untreated transgenic mice, nor nontransgenic mice at final dilutions of serum down to 1:16, indicating that transgenic mice did not spontaneously generate an antibody reaction to A β .

Immunization with A β caused a modest reduction in A β deposits in the frontal cortex, with a significant reduction in the Congo-red-stained area of APP+PS1 mice, and a significant reduction in the A β -immunostained area of APP mice (Fig. 2 and Fig. 3). Reductions of a similar extent were found in hippocampus. We also quantified immunostaining using A β 40- and A β 42-specific antisera, both of which exhibited the same modest reductions found in total A β immunostaining. We suspect that, with a larger sample size, statistically significant partial reductions would be found in all these measures consistent with other recent reports^{9–11}. In general, the percentage reduction in A β deposition was greater in the APP mice than the APP+PS1 mice. The absolute reductions were greater, however, in the doubly transgenic animals. The APP+PS1 mice already had substantial A β deposits by the time vaccinations were initiated¹². Further studies will test whether beginning vaccinations at an earlier age, or combining vaccination with other A β -lowering treatments, will result in more complete protection from A β deposition, and improve the cognitive performance of 15-months-old transgenic mice even further.

Our most important finding here is that A β vaccination protects

transgenic mice from developing memory deficits compared with KLH-immunized (control) transgenic mice. But how important is the learning paradigm in discerning these differences. We have found that in using the reference-memory version of the water maze, mice of this age (15.5 months) have deficits in escape latency, but not retention on the probe trial³. Thus, the more demanding working-memory version of the water-maze task may be essential to detect such differences. Similarly, a spatial task would require intact function of hippocampal and, to a lesser extent, cortical structures, the locations where plaques accumulate earliest and to the greatest extent in these mice^{12–14}.

This vaccination-associated protection from memory impairment occurs in the presence of reduced, but still substantial A β deposits. The mechanism by which immunization with A β blocks learning and memory deficits is not understood. One possibility is that the antibodies neutralize A β in some restricted compartment or deplete a non-deposited form of A β (for example, a soluble form) that is responsible for the memory loss observed. Recently, soluble A β has been proposed as the cause of synapse loss in APP transgenic mice, as some transgenic lines develop reductions in synaptophysin immunoreactivity in dentate gyrus without developing A β deposits¹⁵. A second possibility is that microglia activated by the inoculations¹ can clear the deposited A β , thereby permitting normal cognitive function. This is not easily reconciled with the relatively modest A β clearance detected, although exhaustive regional analyses have yet to be completed. Perhaps even mice that have already developed extensive brain pathology and memory deficits can benefit from vaccinations given later in life. In view of the absence of adverse effects on behaviour and brain functioning, and the protection of memory functions by the A β vaccines, we strongly recommend testing of this and related approaches for the treatment and prevention of Alzheimer's disease. □

Methods

Vaccination protocols

Mice were obtained by breeding Tg 2576 APP transgenic mice¹⁶ with PS1 line 5.1 transgenic mice¹⁷, resulting in nontransgenic, APP, APP+PS1 and PS1 transgenic mice as described by us previously^{12,13}. Human A β 1–42 peptide (Bachem) was suspended in pyrogen-free Type I water at 2.2 mg ml^{-1} then mixed with $10 \times \text{PBS}$ to yield $1 \times \text{PBS}$ and incubated overnight at 37°C . Control mice were injected with KLH that was prepared in the same manner. The antigen suspension was mixed 1:1 with Freund's complete adjuvant and $100 \mu\text{g}$ A β injected subcutaneously by an experimenter who had no role in the behavioural testing. A boost of the same material (prepared freshly) was made in incomplete Freund's at two weeks and injected once monthly for the next three months. Subsequent monthly boosts were made in mineral oil. Mice were vaccinated, beginning at 7.5 months of age. The sample size of each group was: 6 (3 female/3 male) nontransgenic mice vaccinated with A β or KLH; 7 (4 female/3 male) transgenic mice vaccinated with A β ; 7 (4 female/3 male) transgenic mice vaccinated with KLH. The first post-vaccination behavioural testing period was started 5 days after the fifth vaccination at 11.5 months of age. The second behavioural testing period was started at 15.5 months of age, one month after the ninth vaccination. Mice were killed at 16 months of age. We note that transgenic and nontransgenic mice were also tested for performance in the radial-arm water maze at 6 months of age (before vaccination) and all mice performed well.

Radial-arm water maze testing

Experimenters were unaware of the experimental conditions of the mice at the time of testing. The maze consisted of a circular pool 1 m in diameter with six swim alleys (arms) 19 cm wide that radiated out from an open central area (40 cm in diameter), with a submerged escape platform located at the end of one of the arms^{3,18}. Spatial cues were present on the walls and ceiling of the testing room. The escape platform was placed in a different arm each day, forcing mice to use working memory to solve the task. Each day, mice were given the opportunity to learn the location of the submerged platform during four consecutive acquisition trials followed 30 min later by a retention trial (trial 5). On each trial, the mouse was started in one arm not containing the platform and allowed to swim for up to one minute to find the escape platform. Upon entering (all four paws within the swim alley) an incorrect arm or failing to select an arm after 20 s, the mouse was gently pulled back to the start arm for that trial and charged an error. All mice spent 30 s on the platform following each trial before beginning the next trial. On subsequent trials that day, the start arm was varied, so the mouse could not simply learn the motor rule 'second arm to the left', but must learn the spatial location of the platform that day. After the fourth trial was completed, the mice were placed in their home cage for 30 min, then returned to the maze and administered the retention trial. The platform was located in the same arm on each trial within a day, and was in a different arm across days. Over 1–2 weeks of

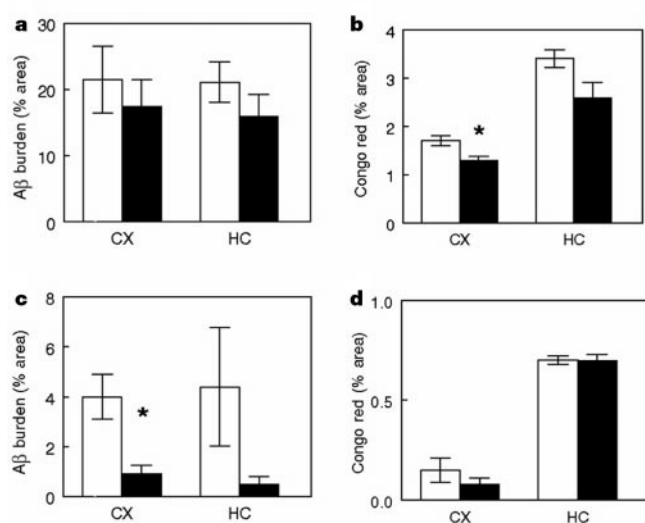


Figure 3 Measurement of amyloid histopathology after A β peptide immunization. **a, b**, Results for the APP+PS1 mice; **c, d**, results for APP-only transgenic mice. A significant reduction in Congo-red staining in frontal cortex was found in APP+PS1 mice vaccinated with A β ($n = 4$) compared to in APP+PS1 mice vaccinated with KLH ($n = 5$; **b**). There was a significant reduction in A β immunostaining in APP-only transgenic mice vaccinated with A β ($n = 3$) compared to in KLH-vaccinated APP mice ($n = 2$; **c**). *, $P < 0.05$; **, $P < 0.01$ by t -test. CX, frontal cortex; HC, hippocampus.

training, control groups gradually improved performance as they learned the procedural aspects of the task, reaching an asymptotic level of 0.5–1 errors on trials 4 and 5. In the experiments presented here mice were trained until the nontransgenic mice reached asymptotic performance: 9 days at 11.5 months or 11 days at 15.5 months. The scores for each mouse on the last two days of testing were averaged and used for statistical analysis. Sensorimotor tests identified no differences among these groups in open field behaviour or string agility testing. As in earlier work, all transgenic mice were impaired on the balance beam, a deficit observed as early as six months of age³, but this deficit was not modified by A β vaccination.

ELISA analysis for serum antibodies

Ninety-six-well Immulon 4HBX (Dynex) micro plates were coated with the A β 1–42 protein (250 ng per well) for 1 h at 37 °C. They were washed four times with 0.45% NaCl + 0.05% Tween-20 (washing buffer, WB). The plates were blocked with 5% non-fat dry milk (NFD) in PBS overnight at 4 °C and washed the following day. Mouse serum was prepared in PBS at an initial dilution of 1:16 and subsequent twofold dilutions were made. All samples were run in duplicate and incubated at 37 °C for 1 h followed by washing 10 times in WB. Plates were blocked a second time with 5% NFD in PBS for 30 min at 37 °C followed by washing five times before the addition of an anti-mouse IgG HRP-conjugate. The secondary antibody was diluted 1:5,000 in PBS and incubated for 1 h at 37 °C. Plates were then washed 10 times in WB and developed with 3,3',5,5'-tetramethylbenzidine substrate (Sigma) in perborate buffer (Sigma). The reaction was stopped with 2 M sulphuric acid. Plates were read spectrophotometrically at 450 nm. The anti-A β 1–42 antibody titre was defined as the reciprocal of the dilution of antisera that produced 50% of the maximum signal detected for that sample.

Histopathology

Mice were overdosed with pentobarbital, perfused with saline and their brains removed. One hemisphere was immersion-fixed in fresh, buffered paraformaldehyde for 24 h. Frozen sections were stained for A β peptides by immunohistochemistry^{13,19} or for Congo red. The area of frontal cortex occupied by stain was measured with a Videometric V150 image analysis system (Oncor) on a Nikon Microphot FX microscope. Stained regions were measured using HSI segmentation by an experimenter unaware of the subject condition. Both stain intensity and area were measured, although only areas are reported here as this is the convention for A β deposits ('amyloid burden'). The results were not qualitatively different when evaluating area, stain intensity or their product (total immunoreactivity¹⁹). Data were collected from equally spaced horizontal sections for both frontal cortex (anterior to the corpus callosum; 12 per mouse) and hippocampus (10 per mouse).

Received 21 August; accepted 7 November 2000.

- Schenk, D. *et al.* Immunization with amyloid-beta attenuates Alzheimer-disease-like pathology in the PDAPP mouse. *Nature* **400**, 173–177 (1999).
- Gordon, M. N. *et al.* Correlation between cognitive deficits and A β deposits in transgenic APP+PS1 mice. *Neurobiol. Aging* (in the press).
- Arendash, G. W. *et al.* Progressive behavioral impairments in transgenic mice carrying both mutant APP and PS1 transgenes. *Brain Res.* (in the press).
- McGeer, E. G. & McGeer, P. L. The importance of inflammatory mechanisms in Alzheimer disease. *Exp. Gerontol.* **33**, 371–378 (1998).
- Rogers, J. *et al.* Inflammation and Alzheimer's disease pathogenesis. *Neurobiol. Aging* **17**, 681–686 (1996).
- Frautschi, S. A. Microglial response to amyloid plaques in APPsw transgenic mice. *Am. J. Pathol.* **152**, 307–317 (1998).
- Masliah, E. *et al.* Comparison of neurodegenerative pathology in transgenic mice overexpressing V717F beta-amyloid precursor protein and Alzheimer's disease. *J. Neurosci.* **16**, 5795–5811 (1996).
- Calhoun, M. E. *et al.* Neuron loss in APP transgenic mice. *Nature* **395**, 755–756 (1998).
- Weiner, H. L. *et al.* Nasal administration of amyloid-beta peptide decreases cerebral amyloid burden in a mouse model of Alzheimer's disease. *Ann. Neurol.* **48**, 567–579 (2000).
- Bard, F. *et al.* Peripherally administered antibodies against amyloid beta-peptide enter the central nervous system and reduce pathology in a mouse model of Alzheimer disease. *Nature Med.* **6**, 916–919 (2000).
- Frenkel, D., Solomon, B. & Benhar, I. Modulation of Alzheimer's beta-amyloid neurotoxicity by site-directed single-chain antibody. *J. Neuroimmunol.* **106**, 23–31 (2000).
- Holcomb, L. *et al.* Accelerated Alzheimer-type phenotype in transgenic mice carrying both mutant amyloid precursor protein and presenilin 1 transgenes. *Nature Med.* **4**, 97–100 (1998).
- Holcomb, L. A. *et al.* Behavioral changes in transgenic mice expressing both amyloid precursor protein and presenilin-1 mutations: Lack of association with amyloid deposits. *Behav. Gen.* **29**, 177–185 (1999).
- McGowan, E. *et al.* Amyloid phenotype characterization of transgenic mice overexpressing both mutant amyloid precursor protein and mutant presenilin 1 transgenes. *Neurobiol. Dis.* **6**, 231–244 (1999).
- Mucke, L. *et al.* High level neuronal expression of A β 1–42 in wild-type human amyloid precursor protein transgenic mice: Synaptotoxicity without plaque formation. *J. Neurosci.* **20**, 4050–4058 (2000).
- Hsiao, K. *et al.* Correlative memory deficits, A β elevation, and amyloid plaques in transgenic mice. *Science* **274**, 99–102 (1996).
- Duff, K. *et al.* Increased amyloid-beta42(43) in brains of mice expressing mutant presenilin 1. *Nature* **383**, 710–713 (1996).
- Diamond, D. M., Park, C. R., Heman, K. L. & Rose, G. M. Exposing rats to a predator impairs spatial working memory in the radial arm water maze. *Hippocampus* **9**, 542–551 (1999).
- Gordon, M. N., Schreier, W. A., Ou, X., Holcomb, L. A. & Morgan, D. G. Exaggerated astrocyte reactivity after nigrostriatal deafferentation in the aged rat. *J. Comp. Neurol.* **388**, 106–119 (1997).

Acknowledgements

We thank K. Hsiao-Ashe for sharing the Tg2576 mouse line. We thank R. Engelman and U. Owens for superb veterinary services in the maintenance of the transgenic colony. We thank I. Johnson for accounting services associated with the project. This work was supported by grants from the National Institute on Aging to M.G. and D.M., from the National Institute of Allergy and Infectious Disease and the National Heart Lung and Blood Institute to K.U., by the Benjamin Research Trust to D.M. and by an award from the University of South Florida Research Foundation.

Correspondence and requests for materials should be addressed to D. M. (e-mail: dmorgan@hsc.usf.edu).

Induction of vanilloid receptor channel activity by protein kinase C

Louis S. Premkumar & Gerard P. Ahern

Department of Pharmacology, Southern Illinois University School of Medicine, Springfield, Illinois 62702, USA

Capsaicin or vanilloid receptors (VRs) participate in the sensation of thermal and inflammatory pain^{1–3}. The cloned (VR1) and native VRs are non-selective cation channels directly activated by harmful heat, extracellular protons and vanilloid compounds^{4–8}. However, considerable attention has been focused on identifying other signalling pathways in VR activation; it is known that VR1 is also expressed in non-sensory tissue^{1,9} and may mediate inflammatory rather than acute thermal pain³. Here we show that activation of protein kinase C (PKC) induces VR1 channel activity at room temperature in the absence of any other agonist. We also observed this effect in native VRs from sensory neurons, and phorbol esters induced a vanilloid-sensitive Ca²⁺ rise in these cells. Moreover, the pro-inflammatory peptide, bradykinin, and the putative endogenous ligand, anandamide, respectively induced and enhanced VR activity, in a PKC-dependent manner. These results suggest that PKC may link a range of stimuli to the activation of VRs.

PKC is a prominent participant in pain signalling. Targeted deletion of PKC- ϵ in mice¹⁰ markedly attenuates thermal- and acid-induced hyperalgesia. In turn, activation of PKC- ϵ potentiates heat-evoked currents in sensory neurons^{11,12}. Further, the algescic peptide, bradykinin, potentiates heat responses^{11,12}, induces depolarization^{13–16}, and evokes secretion^{17–19} from vanilloid-sensitive neurons in a PKC-dependent manner. However, the molecular targets for these effects have not yet been clearly identified. We therefore investigated whether these actions of PKC are mediated by VRs. Rat VR1 was expressed in *Xenopus laevis* oocytes and studied using a two-electrode voltage clamp technique. Treatment with 12-O-tetradecanoylphorbol-13-acetate (TPA) to activate endogenous PKC increased the amplitude of currents evoked by capsaicin (Fig. 1a, c), anandamide (Fig. 1b, c) and protons (extracellular pH 5; data not shown). In addition, TPA by itself produced a slowly developing current (Fig. 1a, b) that was not observed in uninjected oocytes ($n = 5$) or oocytes expressing the NMDA (N-methyl D-aspartate) receptor ($n = 8$). These actions were probably mediated by PKC because no responses were elicited by the inactive TPA analogue, 4 α -phorbol ($n = 4$), and responses to TPA were inhibited by the selective PKC inhibitor²⁰, bisindolylmaleimide (BIM, 200 nM, Fig. 1c).

Next, we examined whether the current induced by TPA alone was mediated by VR1. In these experiments VR1-expressing oocytes were treated separately with either TPA or capsaicin, to avoid cross-sensitization. Figure 1d shows the response of a TPA-treated oocyte to a series of depolarizing pulses from –80 mV to +80 mV. Outwardly rectified currents were evoked that were similar to those

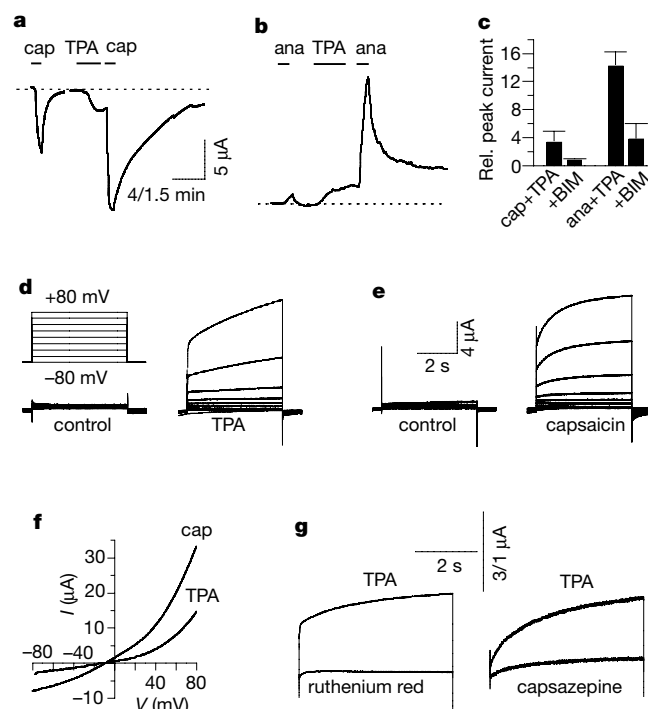


Figure 1 Phorbol ester activates VR1-mediated currents in oocytes. Pretreatment with 500 nM TPA enhanced the current evoked by **a**, 500 nM capsaicin (-40 mV), and **b**, $10 \mu\text{M}$ anandamide ($+30$ mV). Note that TPA alone produced a current. **c**, Mean potentiation of responses evoked by capsaicin ($n = 4$, $P < 0.01$) and anandamide ($n = 7$, $P < 0.001$) and block by 200 nM BIM (capsaicin $n = 3$, $P < 0.05$; anandamide, $n = 5$,

$P < 0.05$). **d**, **e**, Superimposed currents elicited by incremental 20 mV depolarizations from -80 mV. **f**, Representative I - V relationships (500-ms ramp) for currents elicited by capsaicin (500 nM) or TPA (500 nM). The reversal potential, E_{REV} , was -10 and -11 mV, respectively. **g**, TPA-evoked currents ($+40$ mV) were blocked by the VR1 inhibitors ruthenium red ($10 \mu\text{M}$) and capsazepine ($20 \mu\text{M}$).

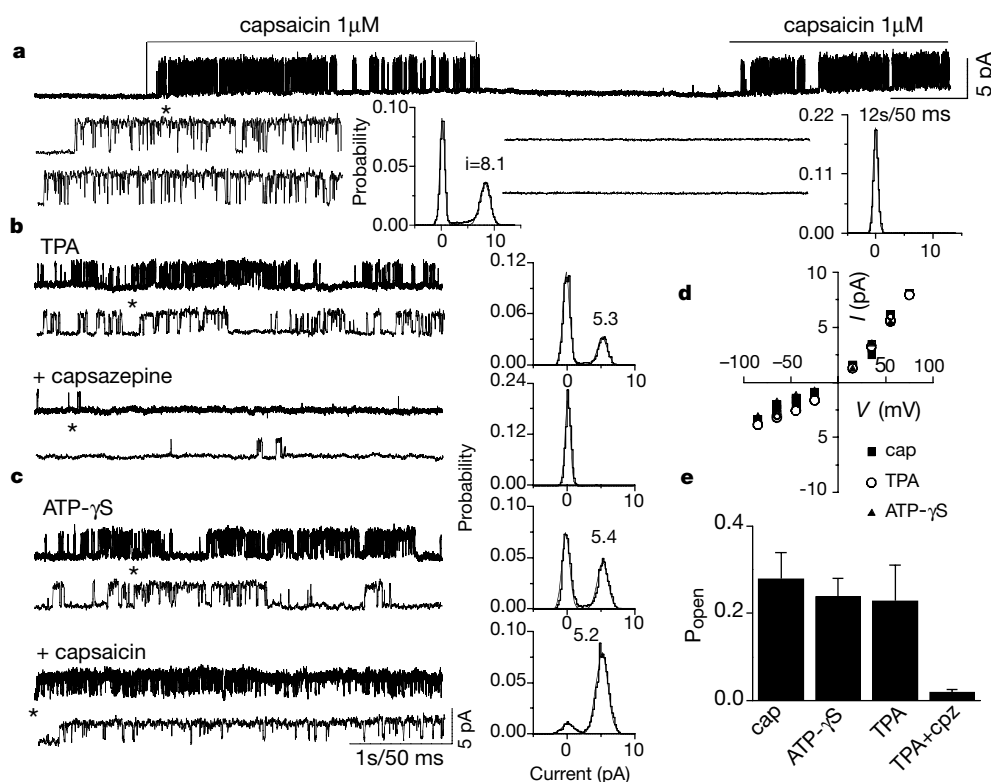


Figure 2 TPA and ATP- γ S induce single channel VR1 activity. Channel activity in outside-out patches from oocytes (asterisk region expanded in lower traces) and gaussian-fitted, all point histograms (right). **a**, Single channel currents ($+75$ mV) evoked by $1 \mu\text{M}$ capsaicin **b**, Channel activity from an oocyte pretreated with TPA (200 nM), and inhibition by capsazepine ($10 \mu\text{M}$). **c**, Channel activity ($P_o = 0.45$) with 1 mM ATP- γ S in the pipette

solution, and further activation ($P_o = 0.89$) by $1 \mu\text{M}$ capsaicin. **d**, I - V relationships of channels activated by capsaicin ($n = 3$), TPA ($n = 2$), and ATP- γ S ($n = 2$). **e**, Mean P_o of channels activated by $1 \mu\text{M}$ capsaicin ($n = 8$), ATP- γ S ($n = 8$), TPA ($n = 5$) and TPA+capsazepine ($n = 3$).

produced by capsaicin (Fig. 1e). These currents were seen in all oocytes tested ($n = 15$), reached a peak within 5–10 min of TPA treatment and were not readily reversible (data not shown). At +40 mV the mean TPA-induced current ($2.9 \pm 0.4 \mu\text{A}$, $n = 6$) was approximately 40% of that produced by 500 nM capsaicin ($7.9 \pm 0.8 \mu\text{A}$, $n = 5$). The current–voltage (I – V) relationships for the currents activated by TPA and those activated by capsaicin were nearly identical (Fig. 1f); both had reversal potentials close to –10 mV (-10.5 ± 0.6 mV, -9.2 ± 0.9 mV, $n = 5$) and exhibited approximately four times greater conductance at +80 mV than at –80 mV. The TPA-activated current also exhibited a similar pharmacological profile to VR1, and was inhibited by capsazepine and ruthenium red (Fig. 1g). These data suggest that activation of PKC leads to the opening of VR1 channels.

To confirm this, we examined the effects of TPA on single VR1 channels in cell-free, outside-out patches. Figure 2a shows that VR1 channel activity was directly related to the application of capsaicin, and little to no activity was evident in the absence of the ligand. In contrast, marked channel activity (Fig. 2b) similar to that evoked by capsaicin was seen in patches excised from oocytes pretreated with TPA (8/12 patches), even though these patches were never exposed to capsaicin. As with capsaicin, the TPA-activated currents were outwardly rectified (Fig. 2d), had a chord conductance of ~ 110 pS (+55 mV, capsaicin, 110 ± 4 pS, $n = 8$; TPA, 107 ± 2 pS, $n = 5$) and a

reversal potential of 3–4 mV (capsaicin, 2.6 ± 2.5 , $n = 3$; TPA, 4.0 ± 0.8 , $n = 2$). Their single channel open probability was 0.23 ± 0.08 ($n = 5$) compared to 0.28 ± 0.06 ($n = 8$) for 1 μM capsaicin (Fig. 2e). Thus, TPA was more potent in excised patches than in intact oocytes; this could be due to the loss of some regulatory factors, including reduced phosphatase activity. In addition, the TPA-activated channels, similar to whole oocyte currents, were sensitive to capsazepine (Fig. 2b; the open probability (P_o) was reduced from 0.24 ± 0.09 to 0.020 ± 0.004 , $n = 3$, $P < 0.05$) and ruthenium red, and could be activated further by capsaicin (data not shown).

If PKC-mediated phosphorylation could activate VR1, then we reasoned that a similar response might be obtained without directly stimulating PKC by inducing irreversible phosphorylation with ATP- γS . Indeed, Fig. 2c and e show that replacing ATP with ATP- γS (1 mM) in the pipette solution produced similar results to TPA (10/14 patches, $P_o = 0.24 \pm 0.04$, $n = 8$). These ATP- γS -activated channels could be blocked by capsazepine (10 μM , $n = 2$), further activated by capsaicin (Fig. 2c), and their current–voltage relationship, reversal potential ($+2.3 \pm 0.9$ mV, $n = 2$) and conductance (103 ± 4 pS, $n = 8$) were similar to capsaicin- and TPA-activated channels (Fig. 2d). In uninjected oocytes ATP- γS produced no response ($n = 9$). These results suggest further that VR activity can be modulated by alterations in phosphatase activity²¹.

Subsequently, we tested whether PKC could also activate native

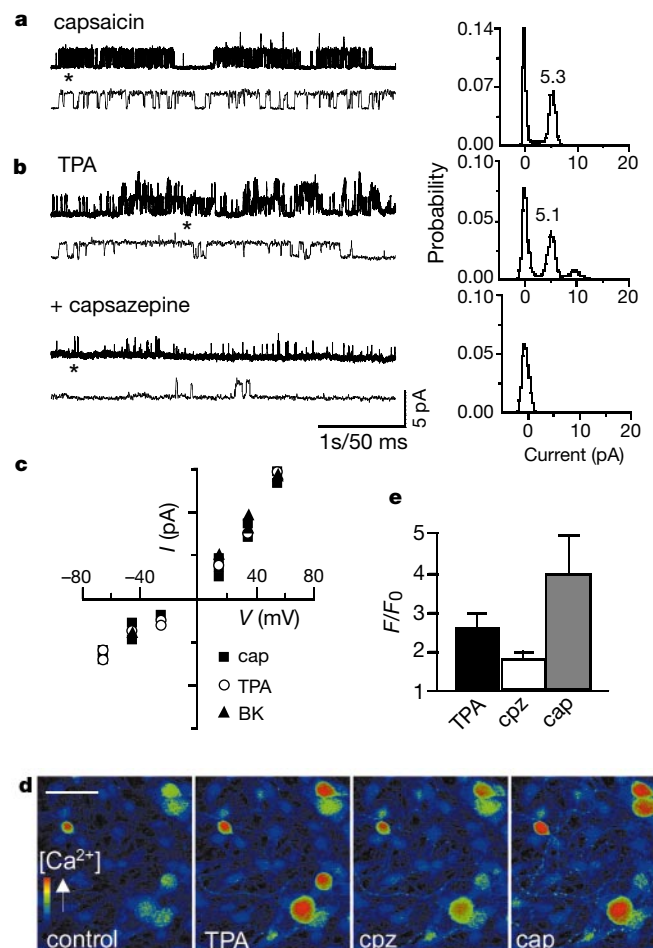


Figure 3 TPA induces VR channel activity and a calcium rise in DRG neurons. **a**, 1 μM capsaicin activated single channel currents in an excised patch. **b**, Channel activity ($P_o = 0.25$) in another patch after TPA (200 nM), and inhibition by 10 μM capsazepine ($P_o = 0.02$). **c**, Superimposable I – V relationships of channels in excised patches activated by capsaicin ($n = 2$), TPA ($n = 3$), and BK ($n = 3$). **d**, Intracellular Ca^{2+} rises

produced by TPA (1 μM) and modulation by sequential bath perfusion of 20 μM capsazepine (cpz) and 5 μM capsaicin (cap). Each panel shows 20 superimposed images collected every 3 s (scale, 50 μm). The mean response of 10 neurons is shown in **e**. (cpz vs TPA, $P < 0.01$; cpz vs cap, $P < 0.05$).

VRs found in dorsal root ganglion (DRG) neurons. We observed much the same response as with cloned receptors. Figure 3a and b show that capsaicin (19/20 patches; $P_o = 0.26 \pm 0.03$, $\gamma = 105 \pm 2$ pS, $n = 6$) and TPA (13/14 patches; $P_o = 0.18 \pm 0.03$, $\gamma = 107 \pm 2$ pS, $n = 5$) induced channel activity with similar properties in these cells. The reversal potentials were 2.7 ± 2.2 mV and 5.3 ± 2.9 mV for capsaicin ($n = 2$) and TPA ($n = 3$), respectively (Fig. 3c), and TPA-induced channel activity exhibited sensitivity to capsazepine (Fig. 3b) and capsaicin (data not shown). As an additional test, to confirm the action of PKC and its specificity to vanilloid-sensitive cells, we performed Ca^{2+} fluorescence imaging of neurons loaded with Fluo-4. Figure 3d and e shows that TPA produced a sustained Ca^{2+} rise in most of the small-diameter neurons ($\Delta F/F_0 = 162 \pm 38\%$, $n = 10$) but had little effect on the surrounding co-cultured glia and fibroblasts. This Ca^{2+} rise, in turn, was inhibited by capsazepine ($\Delta F/F_0 = 79 \pm 20\%$), and a subsequent rise could then be elicited in most of these cells by capsaicin ($\Delta F/F_0 = 298 \pm 99\%$). From three similar experiments we found a tight correlation between sensitivity to TPA and sensitivity to capsaicin (23/25 TPA-responding cells were capsaicin-sensitive; 7/8 non-responding cells were capsaicin-insensitive). Thus, the TPA-induced Ca^{2+} rise occurred mostly in vanilloid-sensitive cells.

As activation of PKC could indeed open native capsaicin receptors, we next considered whether the same pathway might underlie endogenous VR signalling. Bradykinin (BK) is an algescic peptide that binds to the type 2 BK (B2) receptor and can excite sensory neurons in a PKC-dependent manner^{11–17,19}. Thus we hypothesized that BK could activate VRs by a B2-receptor-mediated activation of PKC. Indeed, Fig. 4a shows that BK (5 μM) activated capsaicin-sensitive channels in cell-attached patches (7 of 12 patches, $P_o = 0.25 \pm 0.06$, $n = 5$). In excised patches, these channels had a conductance of 105 ± 3 pS ($n = 5$) and their current–voltage relationship could be superimposed on the capsaicin and TPA data (Fig. 3c). In contrast, BK produced little activity after pretreatment with BIM (200 nM, 1 of 12 patches, $P_o < 0.001$). To explore further the requirement of PKC we studied BK-evoked responses in outside-out

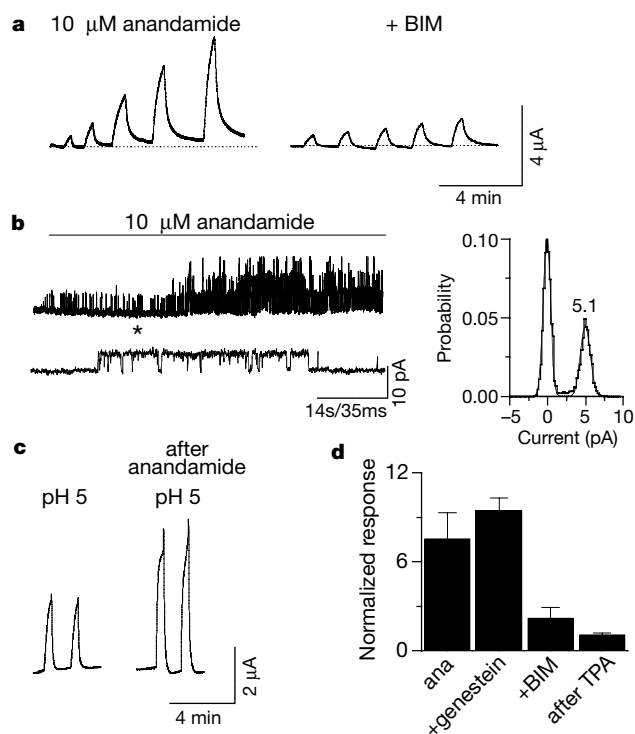


Figure 5 Enhancement of anandamide-evoked VR1 responses via PKC. **a**, Repeated application of 10 μM anandamide evoked progressively larger VR1 currents in oocytes (left, +30 mV). The increase was inhibited by 1 μM BIM (right, +30 mV). **b**, Anandamide (10 μM) produced a similar progressive increase in single channel activity in an outside-out patch (+55 mV). **c**, Anandamide treatment (50 μM , 5 min) enhanced proton-evoked (pH 5) currents consistent with PKC-mediated phosphorylation of VR1. **d**, Mean potentiation of anandamide-evoked current (fifth application); control ($n = 7$), genestein (40 μM , $n = 3$), BIM (1 μM , $n = 5$), TPA (1 μM , $n = 4$).

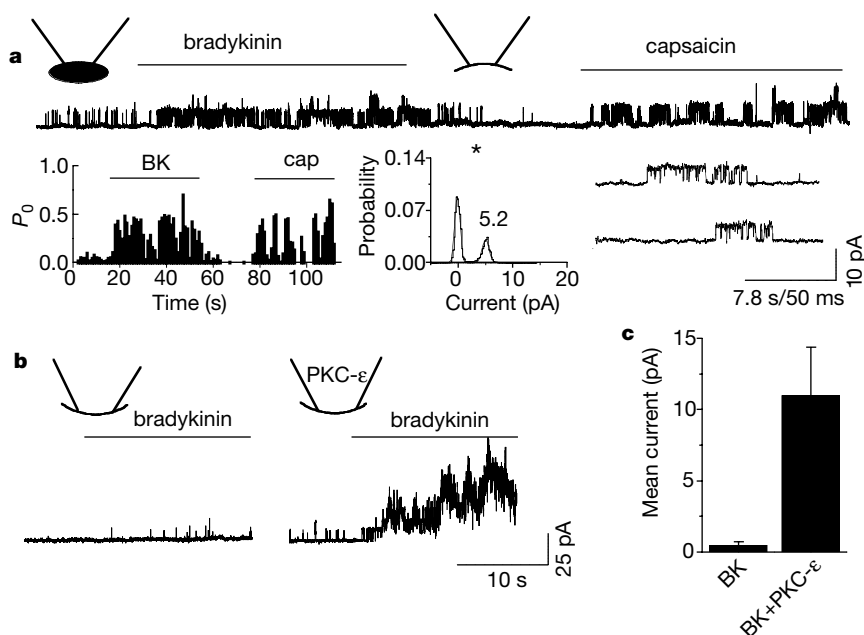


Figure 4 Bradykinin activates VRs via PKC. **a**, Bradykinin (5 μM) activated channels in a cell-attached patch (pipette potential -55 mV). Activity declined after formation of an inside-out patch (end of BK application) and subsequently increased again after application of 1 μM capsaicin. The histogram (left) shows the P_o versus time (1-s bins) for

the entire experiment. The all-points histogram (centre) was constructed from data shortly after patch excision (shown by asterisk). **b**, Activation of VRs by bradykinin in outside-out patches, control (left) and PKC- ϵ (5 $\mu\text{g ml}^{-1}$) in pipette solution (right). **c**, Mean integral current (I) from patches under stated conditions ($n = 5$ in both).

patches. In this configuration marked responses ($P_o = 0.53 \pm 0.07$, $n = 5$) were only seen when PKC- ϵ was included in the pipette solution (Fig. 4b, c). These results indicate that BK activates VRs and that this is mediated through PKC.

Next we considered a modulatory role of the lipid metabolite anandamide in PKC–VR signalling. Anandamide can directly gate VRs and this action is implicated in the regulation of vascular tone^{22,23}. However, anandamide is a weak agonist *in vitro*, and this has prompted a questioning of its biological role at VRs. One resolution of this controversy is that anandamide acts more potently on the phosphorylated VR, as we have now demonstrated (Fig. 1b, c). In the course of these experiments we found that anandamide application itself sensitized VR responses. Figure 5a shows that repetitive stimulation of VR1-expressing oocytes with 10 μ M anandamide produced a progressive increase in the evoked response culminating in a near eight times increase after five applications (Fig. 5d, $n = 7$, $P < 0.05$). We also observed a similar trend in single-channel recordings (Fig. 5b, $n = 5$). In contrast, this potentiation was significantly inhibited by pretreatment with BIM (Fig. 5a, d, $n = 5$, $P < 0.05$) or TPA (Fig. 5d, $n = 4$, $P < 0.05$), but not by the tyrosine kinase inhibitor, genestein (Fig. 5d, $n = 3$). In addition, Fig. 5c shows that proton evoked currents (pH 5) were enhanced (2.4 ± 0.3 fold, $n = 5$) after anandamide treatment, consistent with phosphorylation of VR1. These results strongly suggest that anandamide stimulates PKC²⁴ leading to an enhancement of agonist-evoked responses.

Our results explain why phorbol esters and BK, by themselves, excite capsaicin-sensitive neurons^{13–19} and why disruption of either the VR1 (ref. 3) or PKC- ϵ ¹⁰ gene inhibits thermal hyperalgesia; presumably PKC- ϵ is required to activate VR1. Because VRs are also expressed in non-sensory tissues, mechanisms other than harmful heat and protons are probably needed to gate these channels. Several endogenous agents including anandamide^{22,23} and lipoxygenase metabolites²⁵ have been proposed as endogenous VR ligands, yet possess weak agonistic activity *in vitro*. Our results indicate that production of these ligands and concomitant VR phosphorylation produce a synergistic response. We note that anandamide sensitizes its own response via this same mechanism and thus becomes a more potent agonist. The fact that maximal VR activity requires a combination of signalling events has intriguing implications for both the sensory and non-sensory roles of these channels. In summary, PKC signalling may represent a generalized mechanism for transducing a variety of stimuli into VR activation. □

Methods

Oocyte electrophysiology

Defolliculated *Xenopus laevis* oocytes were injected with 30–70 nl ($1 \mu\text{g} \mu\text{l}^{-1}$) of VR1 cRNA²⁶. Double electrode voltage clamp was performed using a Warner amplifier (Warner Instruments, OC725C) with 100% d.c. gain. Quantitative measurements were made at +40 mV or less where currents were smaller than 10 μ A to ensure adequate voltage clamp. All the experiments were performed at 21–23 °C. Oocytes were placed in a Perspex chamber and continuously superfused (5–10 ml min⁻¹) with Ca²⁺-free Ringer solution containing (in mM): 100 NaCl, 2.5 KCl, 5 HEPES, 1.5 EGTA, pH 7.35. Ca²⁺-free conditions were used to minimize VR1 tachyphylaxis and contamination from Ca²⁺-activated Cl⁻ currents. Electrodes were filled with 3 M KCl and had resistances of 0.5–1 M Ω . *I*–*V* relationships were measured using incremental 20-mV pulses of 5-s duration from a holding potential of –80 to +80 mV; or by 500-ms voltage ramps from –80 mV to 80 mV from a holding potential of 0 mV. Unless otherwise indicated, leak currents measured under control conditions were subtracted from agonist-induced currents. For single channel recording the bathing solution contained (in mM): 100 NaCl, 2.5 KCl, 5 HEPES, 1.5 EGTA, pH 7.3. The patch pipettes were made from glass capillaries (Drummond, Microcaps), coated with sylgard (Dow Corning) and filled with a solution that contained (in mM): 90 Na-gluconate, 10 NaCl, 10 BAPTA, 1 MgCl₂, 10 HEPES, 2 K₂ATP, 0.25 GTP, pH 7.3.

DRG electrophysiology

Dorsal root ganglia were isolated from embryonic day 18 (E18) rats triturated and cultured in Neurobasal/B-27 (Life Technologies)²⁷ +10% fetal bovine serum (FBS) on poly-D-lysine-coated glass coverslips. This procedure generated clusters of neurons interspersed among a layer of glia and fibroblasts. Cells were used after 5–10 d in culture. For excised

outside-out patches, the bath solution contained (in mM): 140 Na-gluconate, 10 NaCl, 1 or 2 MgCl₂, 10 EGTA, 10 HEPES, pH 7.3 and the pipette solution contained (in mM): 140 Na-gluconate, 10 NaCl, 1 or 2 MgCl₂, 10 EGTA, 10 HEPES, 2 K₂ATP pH 7.3. For cell-attached recording external Na-gluconate was replaced with K-gluconate to neutralize membrane potential. Patches that showed activity at 0 mV, indicating the presence of K⁺ channels, were discarded. Agar-bridge electrodes were used to avoid changes in junction potential. The currents were recorded using a Warner Instruments patch clamp amplifier (PC505A) and the data were digitized at 94 kHz (VR-10B, Instrutech Corp.) and stored on video tape.

For analysis the data were filtered at 2.5 kHz (–3 db frequency with an 8-pole low-pass Bessel filter, Warner Instruments, LPF8) and digitized at 5 kHz. Data analyses were performed on continuous stretches greater than 20 s from patches that apparently contained one or two channels. Single channel current amplitude and P_o were calculated from all point amplitude histograms fitted with gaussian functions (Microcal origin). For multiple channel patches mean P_o was measured as NP_o divided by N , where N is the number of channels. Chord conductance was measured at +55 mV and reversal potentials were estimated by extrapolating linear fits to current versus voltage data between +20 and +60 mV. Bradykinin was applied by a perfusion system that has a solution exchange time constant of < 100 ms.

Ca²⁺ imaging

Cells were incubated with 5 μ M Fluo-4AM (Molecular Probes) for 20 min at 25 °C. Fluorescence measurements were made with a Fluoview Olympus confocal microscope. Fluo-4 was excited at 488 nm, and emitted fluorescence was filtered with a 535 ± 25 nm bandpass filter and read into a computer running Fluoview software.

Capsaicin, capsaizapine, TPA, bradykinin, anandamide and ruthenium red were obtained from Sigma. 4 α -phorbol and bisindolylmaleimide were obtained from Calbiochem. PKC- ϵ (human) was obtained from PanVera. Data are given as mean $\pm$ s.e.m., and statistical significance was evaluated using Student's *t*-test.

Received 4 September; accepted 4 October 2000.

- Szallasi, A. & Blumberg, P. M. Vanilloid (Capsaicin) receptors and mechanisms. *Pharmacol. Rev.* **51**, 159–212 (1999).
- Caterina, M. J. *et al.* Impaired nociception and pain sensation in mice lacking the capsaicin receptor. *Science* **288**, 306–313 (2000).
- Davis, J. B. *et al.* Vanilloid receptor-1 is essential for inflammatory thermal hyperalgesia. *Nature* **405**, 183–187 (2000).
- Wood, J. N. *et al.* Capsaicin-induced ion fluxes in dorsal root ganglion cells in culture. *J. Neurosci.* **8**, 3208–3220 (1988).
- Dray, A., Forbes, C. A. & Burgess, G. M. Ruthenium red blocks the capsaicin-induced increase in intracellular calcium and activation of membrane currents in sensory neurones as well as the activation of peripheral nociceptors in vitro. *Neurosci. Lett.* **110**, 52–59 (1990).
- Oh, U., Hwang, S. W. & Kim, D. Capsaicin activates a nonselective cation channel in cultured neonatal rat dorsal root ganglion neurons. *J. Neurosci.* **16**, 1659–1667 (1996).
- Caterina, M. J. *et al.* The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389**, 816–824 (1997).
- Tominaga, M. *et al.* The cloned capsaicin receptor integrates multiple pain-producing stimuli. *Neuron* **21**, 531–543 (1998).
- Mezey, E. *et al.* Distribution of mRNA for vanilloid receptor subtype 1 (VR1), and VR1-like immunoreactivity, in the central nervous system of the rat and human. *Proc. Natl Acad. Sci. USA* **97**, 3655–3660 (2000).
- Khasar, S. G. *et al.* A novel nociceptor signaling pathway revealed in protein kinase C epsilon mutant mice. *Neuron* **24**, 253–260 (1999).
- Cesare, P. & McNaughton, P. A novel heat-activated current in nociceptive neurons and its sensitization by bradykinin. *Proc. Natl Acad. Sci. USA* **93**, 15435–15439 (1996).
- Cesare, P., Dekker, L. V., Sardini, A., Parker, P. J. & McNaughton, P. A. Specific involvement of PKC-epsilon in sensitization of the neuronal response to painful heat. *Neuron* **23**, 617–624 (1999).
- Mizumura, K., Koda, H. & Kumazawa, T. Evidence that protein kinase C activation is involved in the excitatory and facilitatory effects of bradykinin on canine visceral nociceptors in vitro. *Neurosci. Lett.* **237**, 29–32 (1997).
- Dunn, P. M. & Rang, H. P. Bradykinin-induced depolarization of primary afferent nerve terminals in the neonatal rat spinal cord in vitro. *Br. J. Pharmacol.* **100**, 656–660 (1990).
- Burgess, G. M., Mullaney, I., McNeill, M., Dunn, P. M. & Rang, H. P. Second messengers involved in the mechanism of action of bradykinin in sensory neurons in culture. *J. Neurosci.* **9**, 3314–3325 (1989).
- Dray, A., Bettaney, J., Forster, P. & Perkins, M. N. Bradykinin-induced stimulation of afferent fibres is mediated through protein kinase C. *Neurosci. Lett.* **91**, 301–307 (1988).
- Barber, L. A. & Vasko, M. R. Activation of protein kinase C augments peptide release from rat sensory neurons. *J. Neurochem.* **67**, 72–80 (1996).
- Frayer, S. M., Barber, L. A. & Vasko, M. R. Activation of protein kinase C enhances peptide release from rat spinal cord slices. *Neurosci. Lett.* **265**, 17–20 (1999).
- Kessler, F., Habelt, C., Averbeck, B., Reeh, P. W. & Kress, M. Heat-induced release of CGRP from isolated rat skin and effects of bradykinin and the protein kinase C activator PMA. *Pain* **83**, 289–295 (1999).
- Toullec, D. *et al.* The bisindolylmaleimide GF 109203X is a potent and selective inhibitor of protein kinase C. *J. Biol. Chem.* **266**, 15771–15781 (1991).
- Docherty, R. J., Yeats, J. C., Bevan, S. & Boddeke, H. W. Inhibition of calcineurin inhibits the desensitization of capsaicin-evoked currents in cultured dorsal root ganglion neurones from adult rats. *Pflügers Arch.* **431**, 828–837 (1996).
- Zygmunt, P. M. *et al.* Vanilloid receptors on sensory nerves mediate the vasodilator action of anandamide. *Nature* **400**, 452–457 (1999).
- Smart, D. *et al.* The endogenous lipid anandamide is a full agonist at the human vanilloid receptor

- (hVR1). *Br. J. Pharmacol.* **129**, 227–230 (2000).
24. De Petrocellis, L., Orlando, P. & Di Marzo, V. Anandamide, an endogenous cannabinomimetic substance, modulates rat brain protein kinase C in vitro. *Biochem. Mol. Biol. Int.* **36**, 1127–1133 (1995).
25. Hwang, S. W. *et al.* Direct activation of capsaicin receptors by products of lipoxygenases: Endogenous capsaicin-like substances. *Proc. Natl Acad. Sci. USA* **97**, 6155–6160 (2000).
26. Methfessel, C. *et al.* Patch clamp measurement on *Xenopus laevis* oocyte: current through endogenous channels and implanted acetylcholine receptor and sodium channels. *Pflügers Arch* **407**, 577–588 (1986).
27. Brewer, G. J., Torricelli, J. R., Evege, E. K. & Price, P. J. Optimized survival of hippocampal neurons in B27-supplemented Neurobasal, a new serum-free medium combination. *J. Neurosci. Res.* **35**, 567–576 (1993).

Acknowledgements

We thank D. Hallam for assistance with the culture of DRG neurons, and M. Caterina and D. Julius for the gift of VR1. This work was supported by the NSF.

Correspondence and requests for materials may be addressed to L.S.P. (lpremkumar@siumed.edu) or G.P.A. (gahern@siumed.edu).

Co-assembly of polycystin-1 and -2 produces unique cation-permeable currents

Kazushige Hanaoka^{*,†}, Feng Qian^{†,‡}, Alessandra Boletta^{†,‡}, Anil K. Bhunia[‡], Klaus Piontek[‡], Leonidas Tsiokas[§], Vikas P. Sukhatme^{||}, William B. Guggino^{*} & Gregory G. Gergino[‡]

^{*} Department of Physiology and [‡] Department of Medicine, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA

[§] Department of Cell Biology and Warren Medical Research Institute, University of Oklahoma Health Science Center, 941 Stanton L. Young Boulevard, Oklahoma City, Oklahoma 73104, USA

^{||} Renal Division, Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Avenue, Boston, Massachusetts 02215, USA

[†] These authors contributed equally to this work

The human kidney is composed of roughly 1.2-million renal tubules that must maintain their tubular structure to function properly. In autosomal dominant polycystic kidney disease (ADPKD) cysts develop from renal tubules and enlarge independently, in a process that ultimately causes renal failure in 50% of affected individuals^{1,2}. Mutations in either *PKD1* or *PKD2* are associated with ADPKD but the function of these genes is unknown. *PKD1* is thought to encode a membrane protein, polycystin-1, involved in cell–cell or cell–matrix interactions^{3–5}, whereas the *PKD2* gene product, polycystin-2, is thought to be a channel protein⁶. Here we show that polycystin-1 and -2 interact to produce new calcium-permeable non-selective cation currents. Neither polycystin-1 nor -2 alone is capable of producing currents. Moreover, disease-associated mutant forms of either polycystin protein that are incapable of heterodimerization do not result in new channel activity. We also show that polycystin-2 is localized in the cell in the absence of polycystin-1, but is translocated to the plasma membrane in its presence. Thus, polycystin-1 and -2 co-assemble at the plasma membrane to produce a new channel and to regulate renal tubular morphology and function.

Genetic studies have determined that polycystic kidney disease is genetically heterogeneous, with most families having mutations of either *PKD1* on chromosome 16 (~85%) or *PKD2* on chromosome 4 (~15%)^{7–12}. *PKD1* encodes polycystin-1, an integral membrane protein of 4,302 amino acids that has an extracellular amino

terminus of about 3,000 residues containing a previously unknown assortment of motifs found in proteins that mediate cell–cell and cell–matrix interactions. This segment is followed by 11 putative transmembrane domains and a short cytoplasmic carboxy terminus that includes a potential tyrosine phosphorylation site and a coiled-coil structure^{3,4,13}. The *PKD2* gene product, polycystin-2, is an integral membrane protein of 968 amino acids with six putative transmembrane regions⁶. Portions of polycystin-2 share similarities in sequence and membrane topology (Fig. 1a) with ion channels, such as the transient receptor potential and voltage-gated Ca²⁺ and

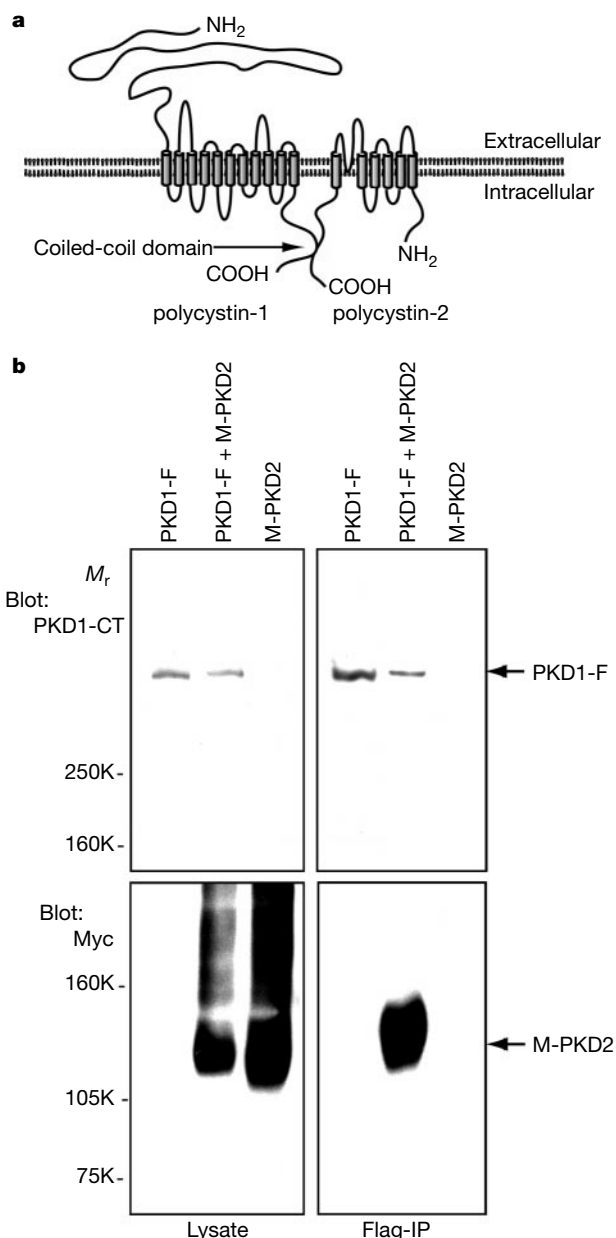


Figure 1 Co-precipitation of polycystin-1 and -2. **a**, Proposed membrane topology and interacting domains (arrow) of polycystin-1 and -2. **b**, CHO cells co-transfected with Flag-polycystin-1 (PKD1-F) and/or Myc-polycystin-2 (M-PKD2). Immunoblots of cell lysates (left) and products immunoprecipitated (IP) with anti-Flag antibody (right) are indicated. Top, CHO cells probed with an anti-PKD1-CT antibody (PKD1-CT), and bottom, with an anti-Myc antibody. PKD1-F is present in the lysate and IP product of samples transfected with *PKD1*. M-PKD2 is present in both of the samples transfected with *PKD2*, but only in the IP product of samples expressing both *PKD1* and *PKD2*.

Na^+ channels, which indicates that polycystin-2 may have similar functional properties^{6,14}. A related polycystin-like protein (PCL), which shares 50% identity and 71% similarity to polycystin-2 (ref. 15), is a Ca^{2+} -regulated non-selective cation channel that is permeable to Ca^{2+} , Na^+ and K^+ when expressed alone in *Xenopus* oocytes¹⁶. Attempts at demonstrating this activity for polycystin-2, however, have been unsuccessful to date. The indistinguishable phenotype resulting from mutations of either *PKD1* or *PKD2* (ref. 1) suggests that their gene products, polycystin-1 and -2, interact to form a functional unit; indeed, the C termini of polycystin-1 and -2 interact through their coiled-coil domains^{13,17}.

To test the hypothesis that interaction with the entire polycystin-1 protein might be necessary for polycystin-2 to function as an ion channel, we co-expressed a unique series of epitope-tagged

full-length complementary DNA constructs of human *PKD1* and *PKD2* in Chinese hamster ovary (CHO) cells. As shown in Fig. 1b, full-length Flag-tagged polycystin-1 (PKD1-F) and Myc-tagged polycystin-2 (M-PKD2) co-immunoprecipitate after co-expression in CHO cells. These data suggest that the full-length proteins are able to associate in mammalian cells to form a heteromeric complex.

We next tested whether co-assembly of polycystin-1 and -2 resulted in new channel activity by measuring whole-cell currents in cells co-expressing both of the polycystin proteins. We found that co-expression of polycystin-1 and -2 (polycystin-1/2) generated time-independent, slightly outwardly rectifying currents that were nearly 20-fold higher than those detected in Mock-transfected cells (0.17 nS for Mock versus 3.2 nS for polycystin-1/2) (Fig. 2a, b). In contrast, whole-cell currents from cells expressing either PKD1-F or M-PKD2 (or an untagged form of polycystin-2) alone were indistinguishable from those observed in Mock-transfected cells (Fig. 2b).

To characterize fully the electrophysiological properties at the whole-cell level that result from formation of the polycystin-1/2

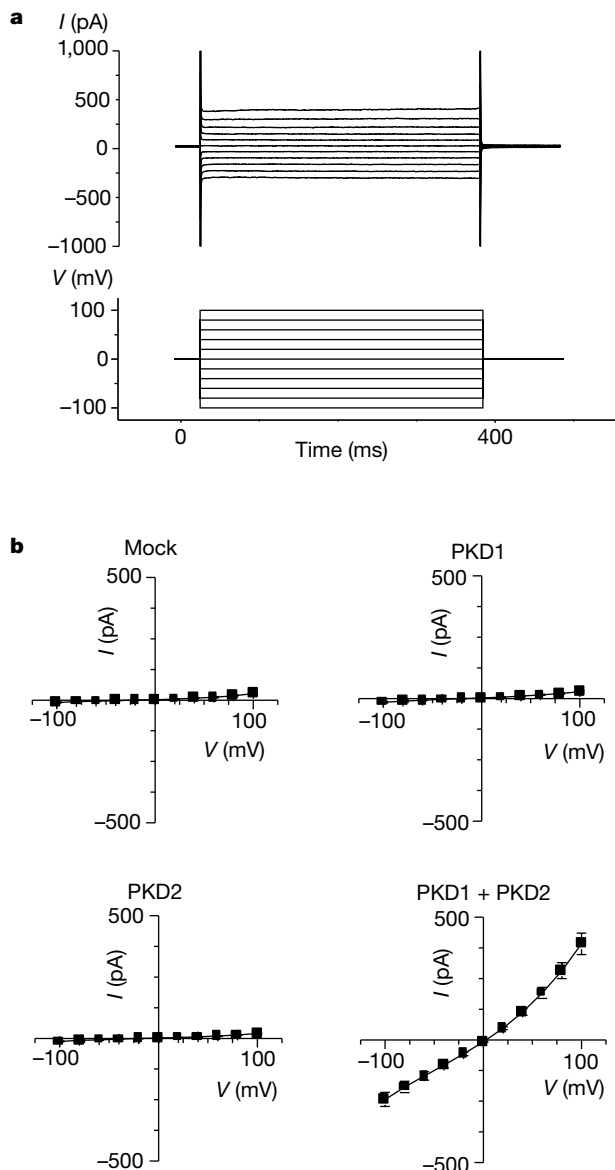


Figure 2 Co-expression of polycystin-1 and -2 results in unique ion-channel activity. **a**, Representative whole-cell currents from cells co-transfected with *PKD1* and *PKD2*. Voltage was clamped between -100 to $+100$ mV (increments of 20 mV for 400 ms). **b**, I - V relationships from cells transfected with Mock ($n = 19$), *PKD1* ($n = 37$), *PKD2* ($n = 34$) or with *PKD1* + *PKD2* ($n = 47$ with increased currents out of 106 total patches). No differences are noted between tagged or wild-type versions of polycystin-2, and the data are combined.

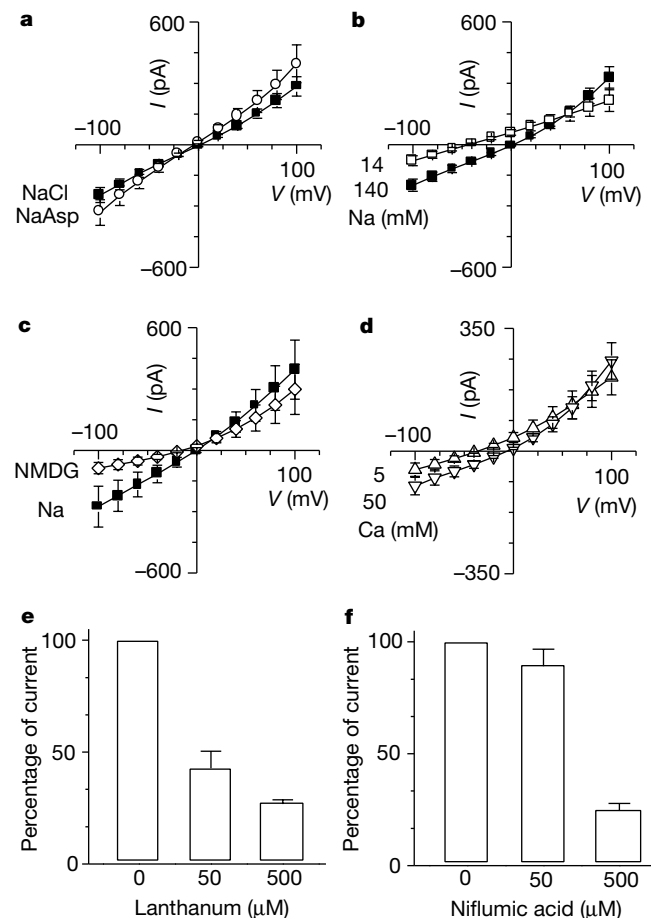


Figure 3 Polycystin-1/2 channel is non-selective for cations. **a-d**, I - V relationships. 140 mM NaCl versus 140 mM sodium aspartate (NaAsp) ($n = 8$) (**a**), 140 mM NaCl versus 140 mM NaCl ($n = 18$) (**b**), 140 mM NaCl versus 140 mM NMDGCl ($n = 9$) (**c**), and 5 mM CaCl_2 versus 50 mM CaCl_2 ($n = 8$) (**d**). E_{rev} (mV) is 2.1 ± 0.6 for 140 mM NaCl, 0.0 ± 1.3 for 140 mM NaAsp ($P > 0.1$), -13.7 ± 3.1 for 140 mM NMDGCl ($P < 0.01$), -6.3 ± 1.5 for 50 mM CaCl_2 ($P < 0.01$), and -36.1 ± 1.8 for 5 mM CaCl_2 ($P < 0.01$). **e, f**, Currents at $+100$ mV compared before and after lanthanum and niflumic acid. 50 or 500 μM of lanthanum reduces currents to $42.8 \pm 7.6\%$ ($n = 6$; $P < 0.01$) and $26.8 \pm 1.9\%$ ($n = 6$; $P < 0.01$), respectively (**e**). 50 μM of niflumic acid has no significant effect ($89.1 \pm 7.0\%$; $n = 9$; $P > 0.05$), whereas 500 μM reduces currents to $24.9 \pm 1.6\%$ ($n = 5$; $P < 0.01$) (**f**).

complex, we tested the ion selectivity using different bath solutions. Replacement of bath Cl^- by aspartate had no effect on whole-cell currents generated by co-expression of polycystin-1 and -2 (Fig. 3a), whereas reduction of Na^+ concentration from 140 to 14 mM caused a significant shift of reversal potential (E_{rev}) (45 mV) toward negative potentials. Reducing the Na^+ concentration also decreased inward currents, suggesting that the currents were carried by cations (Fig. 3b). Replacement of bath Na^+ by *N*-methyl-D-glucamine (NMDG) shifted E_{rev} by 15 mV toward negative potentials, indicating a reduced permeability of the channel to NMDG (Fig. 3c). Thus, the monovalent cation permeability ratio (P) of the polycystin-1/2 channel is $P_{\text{Na}}/P_{\text{Cs}}/P_{\text{NMDG}} = 1.12/1/0.57$.

Similar to the PCL channel¹⁶, the channel resulting from co-expression of polycystin-1 and -2 also passes Ca^{2+} (Fig. 3d). Inward currents were observed in PKD1/2 transfected cells when Ca^{2+} was used as the only permeable cation in the bath solution, whereas no Ca^{2+} currents were observed in mock-transfected cells (data not shown). E_{rev} of the whole-cell Ca^{2+} currents shifted by 30 mV toward negative potentials when the Ca^{2+} concentration in the bath solution was reduced from 50 to 5 mM. This shift in E_{rev} is consistent with that predicted from the Nernst equation for a calcium-permeable channel. The Ca^{2+} selectivity calculated from the E_{rev} (see Methods) measured in 5 mM Ca^{2+} in the bath solution predicts that the Ca^{2+} permeability is about six-times higher than that for monovalent cations such as Na^+ and Cs^+ . Extracellular Ca^{2+} inhibited whole-cell currents generated by monovalent cations. Removing extracellular Ca^{2+} from the bath solution increased whole-cell

currents from -206.4 ± 35.0 pA to -324.2 ± 44.4 pA at -100 mV ($P < 0.01$, $n = 17$); and from 317.5 ± 49.4 pA to 418.5 ± 64.1 pA at $+100$ mV ($P < 0.01$, $n = 17$). The effect of inhibition by extracellular Ca^{2+} was detected at all of the voltages tested (from -100 mV to $+100$ mV), which indicates that Ca^{2+} and monovalent cations may compete for the channel pore. Lanthanum, an inhibitor of non-selective cation channels¹⁸, blocked the currents induced by polycystin-1/2 co-expression in a dose-dependent manner

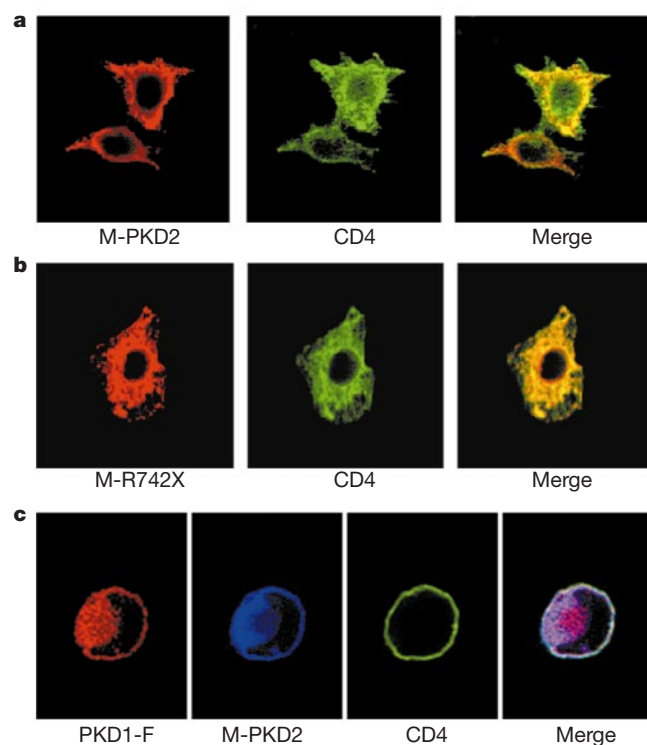


Figure 4 Subcellular localization of CD4, polycystin-1, polycystin-2 and mutant polycystin-2 (R742X) in CHO cells by confocal imaging. **a**, Cells co-transfected with M-PKD2 (red) and CD4 (green). In the absence of polycystin-1, M-PKD2 is restricted to intracellular compartments. Identical results were obtained for three independent transfections (100 cells sampled). **b**, Cells co-transfected with M-R742X (red) and CD4 (green). Almost complete overlap of signal was observed for the two proteins including localization at the surface of the cell. **c**, Cells co-transfected with M-PKD2 (blue), CD4 (green) and PKD1-F (red). Unfixed cells were prestained with anti-CD4. The other antigens were stained after fixing. The three proteins colocalize at the surface of the cell.

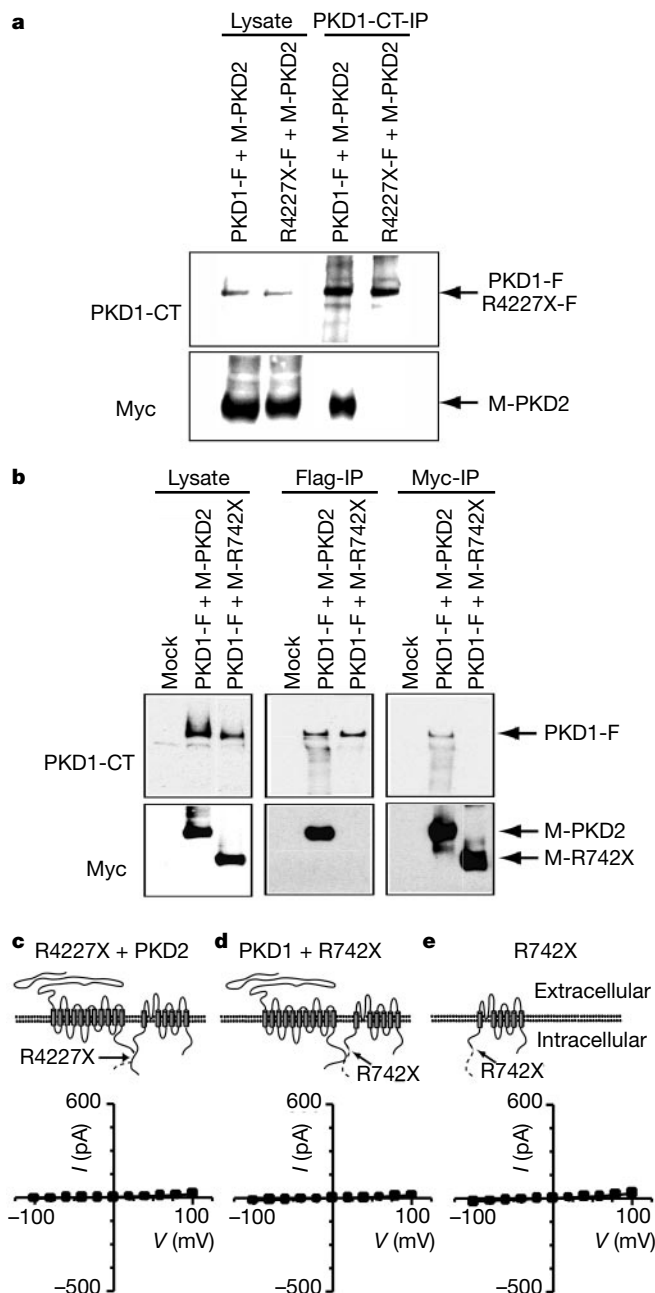


Figure 5 Disease-causing mutations disrupt complex assembly and channel activity. **a**, Immunoblot of PKD1-F and mutant polycystin-1 (R4227X-F) co-expressed with M-PKD2. Lanes 1 and 2 (from left) show equal expression of PKD1-F, R4227X-F and M-PKD2. The R4227X-F mutant has greatly reduced binding to M-PKD2. **b**, Immunoblot of PKD1-F co-expressed with M-PKD2 and mutant polycystin-2 (M-R742X). Equal levels of co-expression are achieved in the cell lysates (left). M-R742X has greatly reduced binding to PKD1-F (middle and right). **c-e**, I - V relationships recorded from R4227X and M-PKD2 ($n = 21$) (**c**), PKD1F and R742X ($n = 25$) (**d**), and R742X alone ($n = 30$) (**e**).

(Fig. 3e). Currents were also inhibited by niflumic acid, an anti-inflammatory drug that inhibits non-selective cation and Cl^- channels¹⁹ (Fig. 3f).

It has been suggested that polycystin-2 is a resident endoplasmic reticulum protein²⁰. Our functional studies, however, suggest that the polycystin-1/2 complex should be present at the plasma membrane if it is directly responsible for the new channel activity. We therefore determined the subcellular localization pattern of polycystin-2 in the absence and presence of polycystin-1. To verify plasma membrane localization, we transfected cells with CD4, a T-lymphocyte-specific cell-surface protein²¹. In patch-clamp experiments, polystyrene beads conjugated with CD4-specific antibodies bound to the extracellular surface of CD4 transfected cells, thereby verifying the plasma membrane localization of CD4. Confocal immunofluorescence images showed that full-length polycystin-2 was localized in the cell and not on the cell surface when expressed alone (Fig. 4a). In contrast, a mutant polycystin-2 protein, R742X, which lacks the terminal 227 amino acids but contains all six transmembrane domains and the possible pore region⁶, colocalized with CD4, suggesting that this mutant form did move to the plasma membrane in the absence of polycystin-1 (Fig. 4b). When full-length polycystin-1 and -2 were co-expressed, CD4 was present at the plasma membrane where it colocalized both with polycystin-1 and -2 (Fig. 4c). The data suggest that polycystin-1 is required for recruitment of polycystin-2 to the plasma membrane.

To show that the channel activity of polycystin-1/2 is dependent on an interaction by means of their respective C-terminal regions, we used naturally occurring disease-causing mutations in both biochemical and electrophysiological assays. A mutant polycystin-1 protein, R4227X, which lacks the terminal 76 amino acids containing the coiled-coil domain of polycystin-1 (ref. 22), was found to have greatly reduced binding to full-length polycystin-2, as measured by co-immunoprecipitation assays (Fig. 5a). Consistent with its inability to form a complex, no currents were generated when R4227X was co-expressed with polycystin-2 (Fig. 5c). Likewise, R742X failed to co-immunoprecipitate with full-length polycystin-1 (Fig. 5b). As predicted by the biochemical analyses, no currents were generated when R742X was expressed alone or co-expressed with full-length polycystin-1 (Fig. 5d, e). As R742X resides in the plasma membrane²⁰ (Fig. 4b), it seems that localization to the plasma membrane of that portion of polycystin-2 predicted to contain the pore is not sufficient to generate channel activity.

Our study provides both biochemical and functional data in support of the hypothesis that polycystin-1 and -2 are interacting partners of a common signalling pathway. We present evidence that co-expression of polycystin-1 and -2 in mammalian cells results in unique channel activity. Although it is possible that another component of the polycystin-1/2 complex forms an ion channel, the similarity between properties of the transmembrane currents measured here and those reported for the PCL channel suggest that polycystin-2 contains the channel-forming pore. Our data also suggest that polycystin-1 is required both to bring polycystin-2 to the plasma membrane and to co-assemble a polycystin complex, enabling polycystin-2 to function as an ion channel. We propose that the normal function of this activity may be to regulate a Ca^{2+} -dependent signalling pathway that controls renal epithelial cell growth and promotes normal tubular morphogenesis and function. The fact that disease-causing mutant forms of these proteins do not generate currents clearly shows that assembly of polycystin-1 and -2 is required for ion-channel function. Although the function of polycystin-1 is unknown, the identical clinical phenotypes associated with mutation of either gene suggest that polycystin-2 is also required for its normal function. Thus, mutations of either polycystin-1 or polycystin-2 disrupt the assembly of the polycystin complex, thereby altering regulation of this pathway and allowing cystogenesis to proceed. □

METHODS

Mammalian cell expression

Full-length cDNAs for human *PKD1* (accession number L33243) and *PKD2* (accession number U50928) were modified to include either a C-terminal Flag epitope (*PKD1*) or an N-terminal Myc epitope (*PKD2*), and cloned into eukaryotic expression vectors pCI-CMV (Promega) and pCDNA3 (Invitrogen), respectively. We used the non-tagged, full-length *PKD2* cDNA in some experiments. The R4227X and R742X mutations of polycystin-1 and polycystin-2, respectively, were introduced using modified oligonucleotide sequences and polymerase chain reaction. For the R4227X mutation, the Flag epitope was inserted before a termination codon and begins at position 4,227. We transiently infected CHO cells with the various expression constructs for *PKD1*, *PKD2* and *CD4* using LipofectAMINE Plus (Life Technologies), according to the manufacturer's instructions. Twenty-four to sixty hours after transfection, some cells were used for electrophysiological studies while the remainder was used for the biochemical studies. We used pCI or pCDNA3 vector for the mock transfection.

Immunoblotting

The cell pellet was resuspended in 1,000 μl of lysis buffer (20 mM sodium phosphate, 150 mM sodium chloride, 10% glycerol, 1 mM EDTA, 0.5% Triton-X 100, pH 7.2 and complete TM protease inhibitor cocktail (Roche) at a concentration of about $1 \mu\text{g} \mu\text{l}^{-1}$) and set on ice for 1 h, then centrifuged at 10,000g at 4 °C for 15 min. The cleared lysate was incubated with 2 μg of 9e10 monoclonal anti-Myc antibody (Roche) or a polyclonal antibody directed against the last 215 amino acids of the C terminus of polycystin-1 (PKD1-CT) at 4 °C for 1 h (ref. 23). We then added 50 μl of protein G sepharose (Pharmacia) to the samples, and allowed them to incubate for a second hour at 4 °C. We washed the immunoprecipitate five times with 1 ml of lysis buffer, then eluted it with 60 μl of laemmli buffer. We used M2-Flag conjugated agarose beads (Sigma) to immunoprecipitate Flag-tagged samples. Fifteen microlitres of the eluted products and an equal volume of whole-cell lysates were subjected to polyacrylamide gel electrophoresis (4% acrylamide for polycystin-1; 6% for polycystin-2), and then analysed using standard western blot protocols (ECL, Amersham).

Solutions

The pipette solution contained (in mM): 15 CsCl, 135 Cs-gluconate, 10 HEPES, 10 glucose, pH 7.4 (adjusted with CsOH). The free Ca^{2+} concentration in pipette solution was fixed to 100 nM by adding 0.2 mM EGTA and 0.12 mM CaCl_2 , as calculated by a programme described elsewhere^{24,25}. In some experiments, 5 mM EGTA and 0.5 mM CaCl_2 were used to obtain the same free Ca^{2+} concentration. The standard bath solution contained (in mM): 140 NaCl, 2.5 CaCl_2 , 10 HEPES, 10 glucose, pH 7.4 (adjusted with NaOH). After the whole-cell configuration was established in standard bath solution, the bath solution was changed to 140 mM NaCl, 14 mM NaCl, 140 mM sodium aspartate, 140 mM NMDGCl, 5 mM CaCl_2 or 50 mM CaCl_2 solutions, to examine the ion selectivity. We fixed the osmolality of the bath solutions by mannitol and we adjusted the pH to 7.4 by using either NaOH, NMDGOH or $\text{Ca}(\text{OH})_2$.

Electrophysiology

Twenty-four to sixty hours after transient transfection of plasmids containing *PKD1*, *PKD2* and *CD4* cDNAs, whole-cell currents were measured from CHO cells grown on glass coverslips. We identified cells expressing CD4 by attachment of CD4 antibody-coated beads (Dynal), and used them for recording. We used fire-polished borosilicate glass pipettes, and applied the pipette solution when the pipette resistance was 5–7 M. We performed perforated whole-cell recordings as described^{26,27} with modification²⁸ using the EPC-7 patch-clamp amplifier (List Electric). We used pClamp7 software (Axon Instruments) for data sampling. Current–voltage (I – V) relationships were obtained at 350 ms of each voltage step and results were shown as mean values $\pm$ s.e.m. We calculated the permeability ratio using the following equation derived from the GHK equation: for monovalent cations, $P_{\text{X}}/P_{\text{Cs}} = \exp(E_{\text{rev}}/RT) [\text{Cs}^+]_i/[\text{X}^+]_o$, where E_{rev} is the reversal potential, F is Faraday's constant, R is the universal gas constant, $[\text{Cs}^+]_i$ and $[\text{X}^+]_o$ are the concentrations of intracellular Cs^+ and extracellular X^+ , respectively, and T is absolute temperature. For measurement of Ca^{2+} permeability, $P_{\text{Ca}}/P_{\text{Cs}} = a_{\text{Cs}}[\text{Cs}^+]_i \{ \exp(E_{\text{rev}}/RT) (1 + \exp(E_{\text{rev}}/RT)) / 4a_{\text{Ca}}[\text{Ca}^{2+}]_o \}$, where $[\text{Ca}^{2+}]$ is the concentration of Ca^{2+} and a is the ion activity coefficient. Assumed ion activity coefficients are 0.75 for Cs^+ , and 0.25 for Ca^{2+} (refs 29 and 30). In all experiments an Ag–agar bridge was used.

Immunofluorescence

CHO cells were fixed with 3% paraformaldehyde and 2% sucrose in PBS (pH 7.6) for 10 min, followed by 5 min of treatment with 1% SDS in 1X TBS to permeabilize membranes. After pre-incubation with 5% BSA/TBS, cells were incubated with anti-Myc antibody for 30 min at 37 °C, washed and then stained with Cy-3-conjugated anti-mouse immunoglobulin (Ig)G antibody (Jackson ImmunoResearch Laboratories) for 30 min. We washed the specimens 4 times for 5 min in TBS, and then incubated them with Fluorescein-conjugated mouse monoclonal anti-CD4 antibody (Calbiochem) for 30 min before imaging. For the triple labelling study, living cells were pre-stained with anti-CD4 antibody at 37 °C for 20 min, washed with PBS, then fixed, permeabilized and pre-blocked as described above. The samples were next treated for 30 min with a blocking solution containing fluorescein isothiocyanate-coupled secondary antibody (anti-murine IgG) at 1:40, washed thoroughly, incubated with anti-PKD1-CT (rabbit) and anti-Myc (mouse) for 30 min, washed again, incubated with rhodamine-conjugated anti-rabbit IgG and Cy-5-conjugated anti-mouse IgG antibody (Jackson ImmunoResearch Laboratories),

washed, and then imaged. We obtained fluorescence images of the cells using a confocal microscope, Noran OZ (Noran Instruments, Middleton, WI).

Received 8 August; accepted 6 November 2000.

1. Gabow, P. A. Autosomal dominant polycystic kidney disease. *N. Engl. J. Med.* **329**, 332–342 (1993).
2. Grantham, J. J. 1992 Homer Smith Award. Fluid secretion, cellular proliferation, and the pathogenesis of renal epithelial cysts. *J. Am. Soc. Nephrol.* **3**, 1841–1857 (1993).
3. The European Polycystic Kidney Disease Consortium. The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. *Cell* **78**, 881–894 (1994).
4. The International Polycystic Kidney Disease Consortium. Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. *Cell* **81**, 289–298 (1995).
5. Hughes, J. et al. The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nature Genet.* **10**, 151–160 (1995).
6. Mochizuki, T. et al. PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science* **272**, 1339–1342 (1996).
7. Reeders, S. T. et al. A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature* **317**, 542–544 (1985).
8. Kimberling, W. J. et al. Linkage heterogeneity of autosomal dominant polycystic kidney disease. *N. Engl. J. Med.* **319**, 913–918 (1988).
9. Peters, D. J. & Sandkuijl, L. A. Genetic heterogeneity of polycystic kidney disease in Europe. *Contrib. Nephrol.* **97**, 128–139 (1992).
10. Harris, P. C. Autosomal dominant polycystic kidney disease: clues to pathogenesis. *Hum. Mol. Genet.* **8**, 1861–1866 (1999).
11. Watnick, T. & Germino, G. G. Molecular basis of autosomal dominant polycystic kidney disease. *Semin. Nephrol.* **19**, 327–343 (1999).
12. Wu, G. & Somlo, S. Molecular genetics and mechanism of autosomal dominant polycystic kidney disease. *Mol. Genet. Metab.* **69**, 1–15 (2000).
13. Qian, F. et al. PKD1 interacts with PKD2 through a probable coiled-coil domain. *Nature Genet.* **16**, 179–183 (1997).
14. Tsiokas, L. et al. Specific association of the gene product of PKD2 with the TRPC1 channel. *Proc. Natl. Acad. Sci. USA* **96**, 3934–3939 (1999).
15. Nomura, H. et al. Identification of PKDL, a novel polycystic kidney disease 2-like gene whose murine homologue is deleted in mice with kidney and retinal defects. *J. Biol. Chem.* **273**, 25967–25973 (1998).
16. Chen, X. Z. et al. Polycystin-L is a calcium-regulated cation channel permeable to calcium ions. *Nature* **401**, 383–386 (1999).
17. Tsiokas, L., Kim, E., Arnold, T., Sukhatme, V. P. & Walz, G. Homo- and heterodimeric interactions between the gene products of PKD1 and PKD2. *Proc. Natl. Acad. Sci. USA* **94**, 6965–6970 (1997).
18. Van Driessche, W., Desmedt, L., De Smet, P. & Simaels, J. Poorly selective cation channels in apical membranes of epithelia. *Experientia Supplementa* **66**, 225–245 (1993).
19. Popp, R., Englert, H. C., Lang, H. J. & Gogelein, H. Inhibitors of nonselective cation channels in cells of the blood-brain barrier. *Experientia Supplementa* **66**, 213–218 (1993).
20. Cai, Y. et al. Identification and characterization of polycystin-2, the PKD2 gene product. *J. Biol. Chem.* **274**, 28557–28565 (1999).
21. Dalgleish, A. G. et al. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **312**, 763–767 (1984).
22. Peral, B. et al. Screening the 3' region of the polycystic kidney disease 1 (PKD1) gene reveals six novel mutations. *Am. J. Hum. Genet.* **58**, 86–96 (1996).
23. Boletta, A. et al. Polycystin-1, the gene product of PKD1, induces resistance to apoptosis and spontaneous tubulogenesis in MDCK cells. *Mol. Cell* **6**, 1267–1273 (2000).
24. Brooks, S. P. & Storey, K. B. Bound and determined: a computer program for making buffers of defined ion concentrations. *Anal. Biochem.* **201**, 119–126 (1992).
25. Hanaoka, K., Wright, J. M., Cheglakov, I. B., Morita, T. & Guggino, W. B. A 59 amino acid insertion increases Ca^{2+} sensitivity of rbsl, a Ca^{2+} -activated K^{+} channel in renal epithelia. *J. Membr. Biol.* **172**, 193–201 (1999).
26. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Arch.* **391**, 85–100 (1981).
27. Hanaoka, K., Devuyst, O., Schwiebert, E. M., Wilson, P. D. & Guggino, W. B. A role for CFTR in human autosomal dominant polycystic kidney disease. *Am. J. Physiol.* **270**, C389–C399 (1996).
28. Fan, J. S. & Palade, P. Perforated patch recording with β -escin. *Pflügers Arch.* **436**, 1021–1023 (1998).
29. Benham, C. D. & Tsien, R. W. A novel receptor-operated Ca^{2+} -permeable channel activated by ATP in smooth muscle. *Nature* **328**, 275–278 (1987).
30. Valera, S. et al. A new class of ligand-gated ion channel defined by P2x receptor for extracellular ATP. *Nature* **371**, 516–519 (1994).

Acknowledgements

This work was supported by NIH grants (to W.G. and G.G.G.), the American Heart Association, the Polycystic Kidney Disease Research Foundation, the National Kidney Foundation and the Arrison Foundation. The authors are members of the Johns Hopkins University PKD Center of Excellence (NIH). G.G.G. is the Blum Scholar of the Johns Hopkins University School of Medicine. We thank J. Neely for immunofluorescence; M. Delannoy for confocal microscopy; A. Hazama for discussion; and S. Agarwal for technical assistance.

Correspondence should be addressed to W.B.G. (e-mail: wguggino@jhmi.edu) or to G.G.G. (e-mail: ggermino@mail.jhmi.edu) and requests for materials should be addressed to G.G.G.

Hypoinsulinaemia, glucose intolerance and diminished β -cell size in S6K1-deficient mice

Mario Pende*, Sara C. Kozma*, Muriel Jaquet†, Viola Oorschot‡, Rémy Burcelin†, Yannick Le Marchand-Brustel§, Judith Klumperman‡, Bernard Thorens† & George Thomas*

* Friedrich Miescher Institute, CH-4058 Basel, Switzerland

† Institute of Pharmacology and Toxicology, CH-1005 Lausanne, Switzerland

‡ Department of Cell Biology, University Medical Center and Research Institute of Biomembranes, 3584CX Utrecht, The Netherlands

§ INSERM E 99-11, Biologie et physiologie de la Nutrition et Signalisation, Faculty of Medicine, 06107 Nice, France

Insulin controls glucose homeostasis by regulating glucose use in peripheral tissues, and its own production and secretion in pancreatic β cells^{1–3}. These responses are largely mediated downstream of the insulin receptor substrates, IRS-1 and IRS-2 (refs 4–8), through distinct signalling pathways. Although a number of effectors of these pathways have been identified, their roles in mediating glucose homeostasis are poorly defined⁹. Here we show that mice deficient for S6 kinase 1, an effector of the phosphatidylinositol-3-OH kinase signalling pathway⁹, are hypoinsulinaemic and glucose intolerant. Whereas insulin resistance is not observed in isolated muscle, such mice exhibit a sharp reduction in glucose-induced insulin secretion and in pancreatic insulin content. This is not due to a lesion in glucose sensing or insulin production, but to a reduction in pancreatic endocrine mass, which is accounted for by a selective decrease in β -cell size. The observed phenotype closely parallels those of preclinical type 2 diabetes mellitus, in which malnutrition-induced hypoinsulinaemia predisposes individuals to glucose intolerance^{10–12}.

Insulin-induced metabolic responses, most notably protein synthesis, are dependent on nutritional state^{13,14}. In turn, increased protein synthesis relies on S6 kinase 1 (S6K1) activation, which regulates the translation of messenger RNAs encoding components of the protein synthetic apparatus¹⁵. Unlike most effectors of insulin action, S6K1 is also sensitive to nutrients, including glucose¹⁶ and amino acids¹³, suggesting that it may serve as an integration point for both anabolic signals. These observations prompted us to analyse glucose homeostasis in S6K1-deficient mice. Recent studies have shown that such mice are viable and fertile, but exhibit a conspicuous reduction in body size during embryogenesis, an effect mostly overcome by adulthood¹⁷. The weak penetrance of the phenotype may arise from increased expression in S6K1-deficient mice of the highly homologous gene, S6K2 (ref. 17).

Although blood glucose levels did not significantly differ in four-month-old fasted wild-type and S6K1-deficient mice (89 ± 14 versus 96 ± 8 mg dl⁻¹, $n = 8$ males; 74 ± 5 versus 71 ± 1 mg dl⁻¹, $n = 22$ females, respectively), the latter revealed a striking 40–50% reduction in circulating insulin levels ($55 \pm 5.3\%$ of wild type for males, $P < 0.01$, $n = 8$; and $61 \pm 6.5\%$ of wild type for females, $P < 0.01$, $n = 17$). To assess whether the reduction in insulin levels affected glucose disposal, we tested fasted mice for glucose tolerance. At early times after glucose injection, S6K1-deficient mice were significantly hyperglycaemic as compared with wild-type mice (Fig. 1a). After the same glucose challenge, S6K1-deficient mice exhibited a severe reduction in the acute insulin secretory response at 2 min (Fig. 1b). To determine whether insulin deficiency was accompanied by insulin resistance in peripheral tissues, we measured glucose use *in vitro* in the slow-twitch non-glycolytic soleus muscle. Both glucose uptake (Fig. 1c) and glycogen synthesis (Fig. 1d) were unaffected *in vitro* in S6K1-deficient mice after

treatment with 50 nM insulin, a concentration sufficient to elicit a maximum response¹⁸. Similar data were obtained for the fast-twitch glycolytic extensor digitorum longus muscle (data not shown). Thus, hypoinsulinaemia and not insulin resistance is the primary lesion predisposing S6K1-deficient mice to hyperglycaemia. The normal fasting glycaemia in such mice, despite low insulin levels, rather suggests that they may be hypersensitive to insulin, which would then dampen the hyperglycaemic response caused by decreased circulating insulin.

Given the reduced body size of S6K1-deficient mice¹⁷, diminished insulin levels might be due to a disproportionate decrease in overall pancreatic size. However, analysis of mouse versus pancreatic weight revealed no difference between wild-type and S6K1-deficient mice (mouse weights 22.6 ± 0.6 g and 21.3 ± 0.9 g versus pancreas weights 230 ± 20 mg and 220 ± 15 mg, respectively, $n = 6$, four-month-old females). This suggested that low levels of blood insulin stem instead from a deficit in the insulin-secreting β cells, which form the islets of Langerhans in the pancreas with the glucagon-, somatostatin- and pancreatic polypeptide (PP)-secreting α , δ , and PP cells, respectively¹⁹. To measure insulin release, we handpicked islets after collagenase digestion of exocrine pancreas. Insulin secretion was assessed by high glucose perfusion, which induced the characteristic first and second phases of insulin release in wild-type islets (Fig. 2a). Although the kinetics of insulin release were similar in S6K1-deficient islets, the amount of insulin secreted per cell was significantly reduced (Fig. 2a). We obtained similar results for Ca^{2+} -dependent insulin release triggered by K^{+} -induced depolarization instead of high glucose (Fig. 2b). These findings imply that glucose-sensing mechanisms, membrane voltage and Ca^{2+} signalling are largely intact in S6K1-deficient islets, whereas secretory potential is compromised. In parallel, analysis of insulin content also revealed significantly decreased levels in S6K1-deficient islets ($67 \pm 8\%$ of insulin content/DNA as compared with wild-type islets, $n = 12$ sets of islets, $P < 0.05$) and pancreas (50.3 ± 3.9 mU for wild type versus 35.1 ± 4.2 mU insulin mg^{-1} of pancreas for S6K1-deficient mice, $n = 5$, $P < 0.05$). Thus, a deficit in both insulin secretion and content of β cells underlies the hypoinsulinaemia of S6K1-deficient mice.

Reduced pancreatic insulin might stem from a decrease in either insulin synthesis or β -cell mass. Analysis of total pancreas showed that insulin transcripts were reduced, whereas glucagon transcripts were increased in S6K1-deficient mice, as compared with wild-type mice (Fig. 3a). The increase in glucagon mRNA may reflect lower circulating insulin, as in insulinopenic rats²⁰. To determine whether the reduction in pancreatic insulin mRNA arose from a defect in transcription, we measured glucose-induced insulin transcription

in size-matched islets from wild-type and S6K1-deficient mice (Fig. 3b). Incubation with a high concentration of glucose resulted in an almost equivalent accumulation of insulin mRNA in both sets of islets, showing that the reduction in pancreatic insulin mRNA levels was not due to a defect in insulin transcription (Fig. 3b).

We next assessed insulin secretory granules and crinophagic activity by electron microscopy of immunogold-labelled ultrathin cryosections of pancreatic islets. Insulin staining in the Golgi stack and secretory apparatus appeared similar in S6K1-deficient and wild-type β cells (Fig. 3c), with morphometric analysis of electron micrographs showing that mature insulin granules were of equivalent diameter (210.2 ± 3.1 versus 214.5 ± 5.8 nm, $n = 60$ granules) and density (14.3 ± 0.3 versus $15.4 \pm 1.8\%$ of β -cell area occupied by granules, $n = 3$ islet sections). In addition, there was no evidence of increased crinophagy as assessed by morphological analysis and by immunostaining with either the endoplasmic reticulum marker protein disulphide isomerase or the lysosomal marker Lamp I (data not shown). Thus, the reduction in pancreatic insulin protein and in S6K1-deficient mice did not appear to arise from a difference in transcription, synthesis or degradation of insulin.

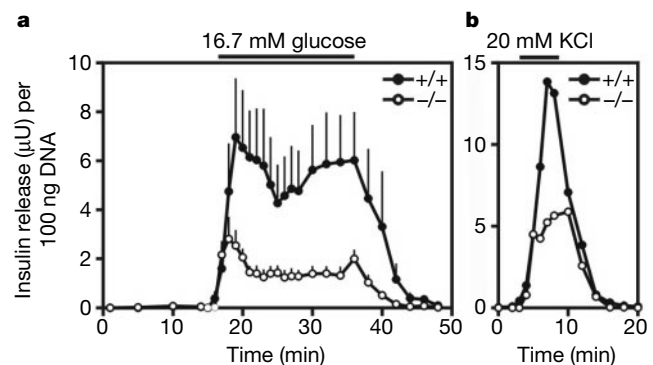


Figure 2 Insulin secretion. **a**, Insulin release was measured from islets of S6K1^{+/+} or S6K1^{-/-} mouse pancreas, during perfusion with a Krebs buffer solution containing 2.8 mM or 16.7 mM glucose for the indicated time. Data represent the average $\pm$ s.e.m. of seven experiments. Total insulin content per islet was $1,290 \pm 461$ μU and 735 ± 123 μU , and DNA content per islet was 9.8 ± 1.0 ng and 11.7 ± 1.1 ng, for S6K1^{+/+} and S6K1^{-/-} islets, respectively. Insulin release was normalized to the DNA content. **b**, Insulin release was stimulated with a Krebs buffer solution containing 20 mM KCl and 2.8 mM glucose. Total insulin content per islet was 710 μU and 648 μU , and DNA content per islet was 11.2 ng and 12.1 ng, for S6K1^{+/+} and S6K1^{-/-} islets, respectively. The result is typical of four independent experiments.

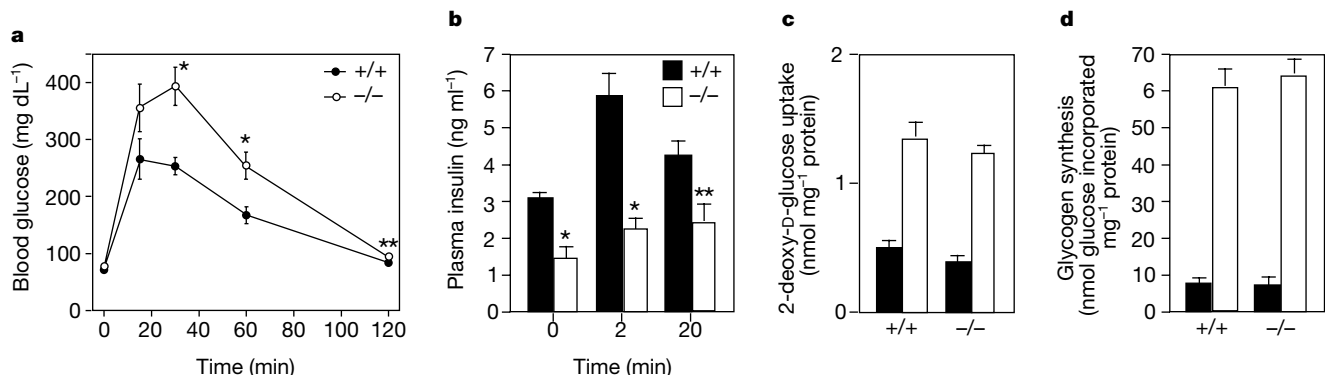


Figure 1 Analysis of glucose tolerance, plasma insulin levels and insulin action. **a**, **b**, Blood glucose (**a**) and plasma insulin (**b**) concentrations before and after intraperitoneal injection of 3 g D-glucose per kg body weight in mice fasted overnight. Data represent the average $\pm$ s.e.m. for seven female mice of each genotype. Mouse weight was 21.5 ± 0.83 and 19.1 ± 0.51 for the S6K1^{+/+} and S6K1^{-/-} mice, respectively. The experiments were repeated four times with similar results in both male and female mice. Asterisk,

$P < 0.005$, two asterisks, $P < 0.05$ versus S6K1^{+/+} mice. **c**, **d**, 2-Deoxy-D-glucose uptake (**c**) and glycogen synthesis (**d**) in isolated soleus muscles of S6K1^{+/+} and S6K1^{-/-} mice. Left and right soleus muscles were incubated in the absence (filled bars) and the presence (open bars) of 50 nM insulin. Data represent the average $\pm$ s.e.m. for five animals of each genotype.

These findings indicated that reduced β -cell mass may be responsible for decreased pancreatic insulin levels in S6K1-deficient mice. Haematoxyline and eosine stained pancreatic sections showed that islets from S6K1-deficient mice were smaller than wild-type islets (Fig. 4a, left panels), with morphometry revealing about a one-third decrease in endocrine mass (Fig. 4b), consistent with reduced pancreatic insulin. Moreover, immunostaining of pancreatic sections with insulin and glucagon antibodies revealed a sharply lower ratio of β -cell to α -cell mass in S6K1-deficient mice (Fig. 4c). As *Drosophila* that are deficient for the fly orthologue of S6K1 are reduced in mass owing to a reduction in cell size²¹, we measured the density of β cells in a constant area within individual islets. Hoechst staining of nuclei revealed that insulin-positive cells were more densely packed in S6K1-deficient islets (Fig. 4a, right), as quantified over a large number of islets (Fig. 4d). The calculated volume of an S6K1-deficient β cell would then be 24% smaller than that of a wild-type cell ($1,059 \mu\text{m}^3$ versus $1,397 \mu\text{m}^3$, respectively). Point-counting morphometry of electron micrographs showed that the effect on cell size was due to a reduction of cytosol, as the cross-sections of nuclei had equivalent diameter whereas the ratio of cytosol to nuclear area was 16% lower in S6K1-deficient than in wild-type β cells (6.5 ± 0.4 versus 7.8 ± 0.4 ; $P < 0.05$, $n = 3$ islet sections).

Decreased cell size was not observed for other endocrine cells, including α and adrenal cells, nor for a number of other cell types (data not shown). Moreover, there was no increased apoptotic activity in β cells of S6K1-deficient mice, as judged by propidium iodide and Hoechst staining, or by electron microscopy, ruling out higher β -cell density owing to increased cell death.

The calculated reduction in β -cell size is roughly equivalent to the reduction in insulin content and endocrine mass. As small β -cell size has a pronounced effect on secretory potential (for example, a 20% reduction in β -cell surface leads to a 35% reduction in insulin

secretion^{22,23}), the reduction in membrane surface would explain the loss in secretory potential. A direct effect of S6K1 on insulin secretion is ruled out by rapamycin, an inhibitor of S6K, which has no immediate effect on insulin secretion in rat islets under the conditions described in Fig. 2a (M. L. McDaniel, personal communication; and ref. 24). Thus, the decrease in both insulin content and secretion appears to be due to reduced β -cell size.

Type 2 diabetes mellitus is a polygenic, multifactorial disease, in which decreased insulin secretion and reduced β -cell mass contribute to its later development¹¹. In cases of protein malnutrition during gestation or childhood, however, these defects in pancreas function are thought to be the primary factors leading to type 2 diabetes later in life¹². Corroborative *in vivo* data for this hypothesis has come from studies in rats, in which a limited period of protein malnutrition leads to glucose intolerance arising from a persistent decrease in β -cell size and insulin secretion, an effect partially attenuated by mild insulin hypersensitivity of peripheral

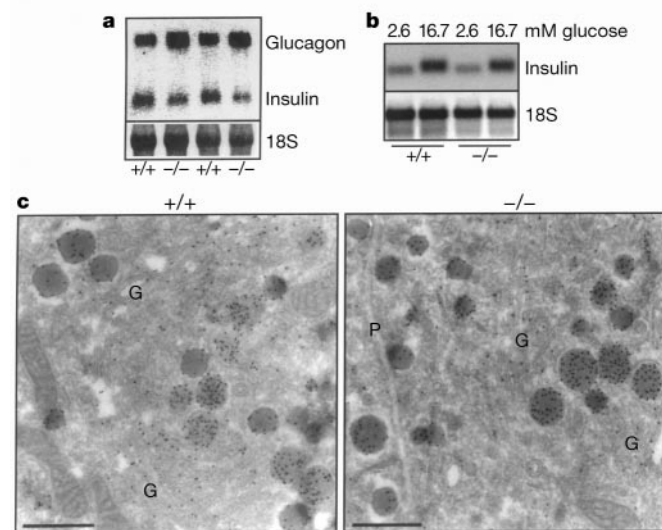


Figure 3 Insulin production. **a**, Total RNA was extracted from S6K1^{+/+} and S6K1^{-/-} mouse pancreas and northern blot analysis was performed to detect the levels of insulin and glucagon mRNAs. 18S ribosomal RNA was visualized by methylene-blue staining of the blot as a control for RNA loading. **b**, Islets from S6K1^{+/+} and S6K1^{-/-} mouse pancreas were incubated overnight in a medium containing 2.8 mM or 16.7 mM glucose. After 18 h, total RNA was extracted and insulin mRNA and 18S rRNA levels were assessed by northern blot analysis. The experiment was repeated two times with similar results. **c**, Ultrathin cryosections of pancreatic β cells of S6K1^{+/+} and S6K1^{-/-} mice. Insulin is readily detectable in the Golgi (G) of both S6K1^{+/+} and S6K1^{-/-} mice, indicating an active synthesis apparatus. Both immature and mature secretory granules (the latter are recognized by the condensed core and electron lucent halo) are morphologically identical in both animals and exhibit similar high labelling densities for insulin. P, plasma membrane. Scale bars, 500 nm.

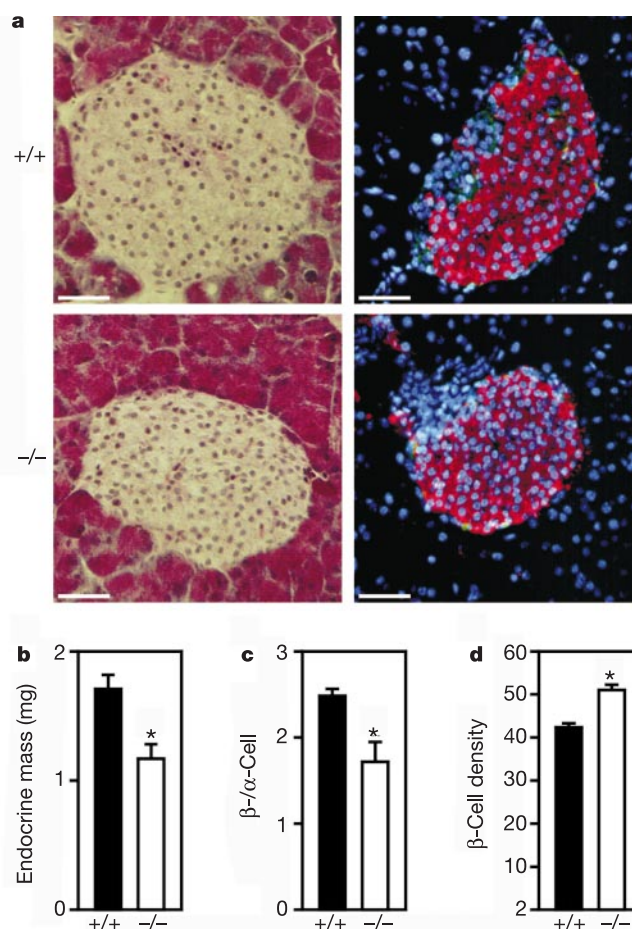


Figure 4 Islet morphology and β -cell mass. **a**, Haematoxyline and eosine stain (left) and immunostain for insulin- and glucagon-containing cells (right) in pancreas sections from S6K1^{+/+} and S6K1^{-/-} mice. Insulin is detected by Alexa Fluor 594-labelled secondary antibody (red), glucagon by fluorescein-labelled secondary antibody (green), and nuclear DNA by Hoechst staining (blue). Scale bar, 50 μm . **b**, Endocrine mass. Data represent average $\pm$ s.e.m. for six female mice of each genotype. Asterisk, $P < 0.02$ versus S6K1^{+/+} mice. Mouse weight was 22.6 ± 0.6 g for S6K1^{+/+} and 21.3 ± 0.9 g for S6K1^{-/-} mice. Pancreatic weight was 230 ± 20 mg for S6K1^{+/+} and 220 ± 15 mg for S6K1^{-/-} mice. **c**, β -Cell to α -cell mass ratio. At least 30 islets from 2–3 pancreas sections were analysed for each animal. Data represent mean $\pm$ s.e.m. for four female mice of each genotype. Asterisk, $P < 0.025$ versus S6K1^{+/+} islets. **d**, β -Cell density was assessed by counting the number of nuclei in a $5,300\text{-}\mu\text{m}^2$ islet area that contained only insulin-positive cells. Values are the mean $\pm$ s.e.m. of 50 determinations done in four female mice of each genotype. Asterisk, $P < 0.001$ versus S6K1^{+/+} β cells.

tissues^{10,23,25}. Consistent with these observations, we find that S6K1-deficient mice are glucose intolerant owing to hypoinsulinaemia and reduced β -cell size. In addition, the growth of S6K1-deficient mice is strongly retarded during embryogenesis¹⁷, and although they largely overcome this deficit in adulthood, they fail to obtain normal levels of β -cell mass, as in human and rat^{10,12}. Unexpectedly, this effect has proved to be selective for β cells, despite the ubiquitous expression of S6K1 and the presence of S6K2 in β cells (data not shown). This may be due to the fact that β -cell growth, like S6K1 activity, is extremely sensitive to mitogens such as insulin and insulin-like growth factors, as well as to nutrients such as glucose and amino acids^{16,26}. Hence, β cells lacking S6K1 may be reduced in size because they fail to efficiently integrate the multiple growth factor and nutrient stimulatory inputs required for β -cell development. Our studies identify S6K1 as a critical effector in controlling glucose homeostasis and indicate that impaired S6K1 function may contribute with other genetic and environmental factors to the development of specific forms of diabetes mellitus. □

Methods

Animals

S6-kinase-deficient mice, previously created by homologous recombination¹⁷, were genotyped as described¹⁷ and kept in a hybrid background derived from the C57Bl/6 and 129/Ola mouse strains. Animals were maintained on a 12-h light/dark cycle and were allowed free access to food. Four-month-old male or female mice were used for experiments.

Glucose tolerance tests and *in vivo* insulin secretion

Mice were fasted for 14 h followed by intraperitoneal injection of glucose (3 g per kg body weight). Whole venous blood was obtained from the tail vein at the indicated time points after the glucose load. Blood glucose levels were measured using a Glucotrend glucometer (Roche). To measure insulin levels at the indicated time during a glucose tolerance test, blood was collected in heparinized microhaematocrit tubes, and centrifuged. Insulin levels were measured in 5–10 μ l of plasma using an ELISA kit for rat insulin (Crystal Chem.).

Glucose uptake and glycogen synthesis in isolated muscles

We measured 2-[³H]deoxyglucose uptake and [³H]glucose incorporation into glycogen in isolated soleus and EDL muscles as described¹⁸.

Insulin release and content of islets

Islets were handpicked after collagenase digestion of pancreas. *In vitro* insulin release from isolated islets was measured using perfusion chambers as described²⁷. We measured insulin content by radioimmunoassay (Linco, St. Louis, MO) in water extracts of isolated islets or acid/ethanol extracts of whole pancreas. DNA content of islets was measured by fluorometric analysis using Hoechst 33258 reagent²⁸. Total RNA was obtained using the acid guanidium thiocyanate/phenol/chloroform extraction²⁷.

Histological analysis

For electron microscopy analysis, we fixed pancreases by injecting at several points 2% formaldehyde and 0.2% glutaraldehyde in 0.1 M phosphate buffer, pH 7.4. Islets of Langerhans were isolated from the pancreatic tissue and either embedded in 10% gelatin for cryosectioning, or postfixed with 1% OsO₄ in 0.1 M cacodylate buffer and embedded in Epon. For immuno-electron microscopy, we prepared ultrathin cryosections and immunolabelled them according to the protein A/gold method²⁹, with a guinea pig anti-insulin antibody (Linco), the monoclonal antibody ID3 against protein disulphide isomerase (a gift from S. Fuller, EMBL, Heidelberg) and the polyclonal antibody 1D4B raised against rat lysosomal enzyme Lamp 1 (a gift from T. August, Developmental Studies Hybridoma Bank, Univ. of Ohio). For quantitation of β -cell cytosol to nuclear ratio, we prepared ultrathin sections after embedding the cells in Epon and stained them with 2% uranyl acetate in distilled water for 30 min at room temperature. At least nine non-overlapping fields of the sections were photographed at a final magnification of $\times 8,500$, and β -cell cytosol and nuclear areas were quantified by point-counting morphometry³⁰. For granule size and density, at least 15 non-overlapping fields of Epon embedded sections were photographed at a final magnification of $\times 41,000$, granule diameter was measured in randomly selected mature granules, and cell area occupied by granules was quantified by point-counting morphometry.

For light microscopy analysis, pancreases were dissected, weighed, fixed overnight in 4% paraformaldehyde and then incubated overnight in 0.6 M sucrose. Four consecutive frozen sections (8 μ m) from five distinct levels of pancreas, 200 μ m apart, were mounted on slides. For the quantification of the endocrine mass, five sections per pancreas were stained with haematoxyline and eosine and photographed at a final magnification of $\times 5.5$. Images were scanned and analysed using the Adobe Photoshop program to quantify the endocrine area versus total pancreas area. This value was multiplied by the weight of the pancreas to get the absolute endocrine mass. For the quantification of β -cell and α -cell mass ratio, sections were immunostained with both rabbit anti-glucagon (Zymed) and guinea pig anti-insulin (Linco) antibodies. Detection was performed using Alexa Fluor

594 (Molecular Probes) and fluorescein secondary antibodies (Cappel). Sections were incubated for 5 min with 0.01% Hoechst to reveal nuclei. All islets in the sections were photographed at a final magnification of $\times 280$, and α -cell and β -cell area was quantified by the point-counting method³⁰.

Received 21 June; accepted 20 September 2000.

- Kahn, B. B. Type 2 diabetes: when insulin secretion fails to compensate for insulin resistance. *Cell* **92**, 593–596 (1998).
- Leibiger, I. B., Leibiger, B., Moede, T. & Berggren, P. O. Exocytosis of insulin promotes insulin gene transcription via the insulin receptor/PI-3 kinase/p70 s6 kinase and CaM kinase pathways. *Mol. Cell* **1**, 933–938 (1998).
- Kulkarni, R. N. *et al.* Tissue-specific knockout of the insulin receptor in pancreatic beta cells creates an insulin secretory defect similar to that in type 2 diabetes. *Cell* **96**, 329–339 (1999).
- Tamemoto, H. *et al.* Insulin resistance and growth retardation in mice lacking insulin receptor substrate-1. *Nature* **372**, 182–186 (1994).
- Araki, E. *et al.* Alternative pathway of insulin signalling in mice with targeted disruption of the IRS-1 gene. *Nature* **372**, 186–190 (1994).
- Withers, D. J. *et al.* Disruption of IRS-2 causes type 2 diabetes in mice. *Nature* **391**, 900–904 (1998).
- Withers, D. J. *et al.* Irs-2 coordinates Igf-1 receptor-mediated beta-cell development and peripheral insulin signalling. *Nature Genet.* **23**, 32–40 (1999).
- Kulkarni, R. N. *et al.* Altered function of insulin receptor substrate-1-deficient mouse islets and cultured beta-cell lines. *J. Clin. Invest.* **104**, R69–R75 (1999).
- Shepherd, P. R., Withers, D. J. & Siddle, K. Phosphoinositide 3-kinase: the key switch mechanism in insulin signalling. *Biochem. J.* **333**, 471–490 (1998); erratum *ibid.* **335**, 711.
- Svenne, I., Borg, L. A., Crace, C. J. & Schnell Landstrom, A. Persistent reduction of pancreatic β -cell mass after a limited period of protein-energy malnutrition in the young rat. *Diabetologia* **35**, 939–945 (1992).
- DeFronzo, R. A. Pathogenesis of type 2 diabetes: metabolic and molecular implications for identifying diabetes genes. *Diabetes Rev.* **5**, 177–269 (1997).
- Phillips, D. I. Insulin resistance as a programmed response to fetal undernutrition. *Diabetologia* **39**, 1119–1122 (1996).
- Hara, K. *et al.* Amino acid sufficiency and mTOR regulate p70 S6 kinase and eIF-4E BP1 through a common effector mechanism. *J. Biol. Chem.* **273**, 14484–14494 (1998).
- Castellino, P., Luzi, L., Simonson, D. C., Haymond, M. & DeFronzo, R. A. Effect of insulin and plasma amino acid concentrations on leucine metabolism in man. Role of substrate availability on estimates of whole body protein synthesis. *J. Clin. Invest.* **80**, 1784–1793 (1987).
- Jefferies, H. B. J. *et al.* Rapamycin suppresses 5'TOP mRNA translation through inhibition of p70^{s6k}. *EMBO J.* **12**, 3693–3704 (1997).
- Hugl, S. R., White, M. F. & Rhodes, C. J. Insulin-like growth factor I (IGF-I)-stimulated pancreatic beta-cell growth is glucose-dependent. Synergistic activation of insulin receptor substrate-mediated signal transduction pathways by glucose and IGF-I in INS-1 cells. *J. Biol. Chem.* **273**, 17771–17779 (1998).
- Shima, H. *et al.* Disruption of the p70^{s6k}/p85^{s6k} gene reveals a small mouse phenotype and a new functional S6 kinase. *EMBO J.* **17**, 6649–6659 (1998).
- Stenbit, A. E. *et al.* Diverse effects of Glut 4 ablation on glucose uptake and glycogen synthesis in red and white skeletal muscle. *J. Clin. Invest.* **98**, 629–634 (1996).
- Slack, J. M. Developmental biology of the pancreas. *Development* **121**, 1569–1580 (1995).
- Dumontel, E. *et al.* Insulin, but not glucose lowering corrects the hyperglucagonemia and increased proglucagon messenger ribonucleic acid levels observed in insulinopenic diabetes. *Endocrinology* **139**, 4540–4546 (1998).
- Montagne, J. *et al.* Drosophila S6 kinase: A regulator of cell size. *Science* **285**, 2126–2129 (1999).
- Giordano, E. *et al.* B-cell size influences glucose-stimulated insulin secretion. *Am. J. Physiol.* **265**, C358–C364 (1993).
- Svenne, I., Crace, C. J. & Jansson, L. Intermittent protein-calorie malnutrition in the young rat causes long-term impairment of the insulin secretory response to glucose *in vitro*. *J. Endocrinol.* **118**, 295–302 (1988).
- Fabian, M. C., Lakey, J. R., Rajotte, R. V. & Kneteman, N. M. The efficacy and toxicity of rapamycin in murine islet transplantation. *In vitro* and *in vivo* studies. *Transplantation* **56**, 1137–1142 (1993).
- Grace, C. J., Svenne, I., Kohn, P. G., Strain, A. J. & Milner, R. D. Protein-energy malnutrition induces changes in insulin sensitivity. *Diabetes Metab.* **16**, 484–491 (1990).
- Svenne, I. Pancreatic beta-cell growth and diabetes mellitus. *Diabetologia* **35**, 193–201 (1992).
- Guillam, M. T. *et al.* Early diabetes and abnormal postnatal pancreatic islet development in mice lacking Glut-2. *Nature Genet.* **17**, 327–330 (1997); errata *ibid.* **17**, 503 (1997), and **88** (1998).
- Hopcroft, D. W., Mason, D. R. & Scott, R. S. Standardization of insulin secretion from pancreatic islets: validation of a DNA assay. *Horm. Metab. Res.* **17**, 559–561 (1985).
- Kuliawat, R., Klumperman, J., Ludwig, T. & Arvan, P. Differential sorting of lysosomal enzymes out of the regulated secretory pathway in pancreatic beta-cells. *J. Cell Biol.* **137**, 595–608 (1997).
- Weibel, E. R. Morphometry of the human lung: the state of the art after two decades. *Bull. Physiopathol. Respir.* **15**, 999–1013 (1979).

Acknowledgements

We thank T. Gremeaux and P. Kopf for expert technical assistance, and M. Rothnie, R. Scriwaneck and T. van Rijn for their help in the rapid and high-quality production of photographic materials. We also thank C. Bendotti for help with statistical analysis. We are grateful to D. Withers, P. Caroni, S. Volarevic, G. Posthuma, J. Slot, H. Geuze, L. Hansen, S. Oldham and W. Krek for reading the manuscript and helpful discussions, and to the members of the Thomas laboratory for support. M.P. is a recipient of a stipend from the EEC and these studies were supported in part by grants from INSERM and Fondation pour la Recherche Médicale to Y.L.-B., from the Swiss National Science Foundation to B.T. and from the EEC, HFSP and Novartis Foundation to G.T.

Correspondence and requests for materials should be addressed to G.T. (e-mail: gthomas@fmi.ch).

Controlled growth factor release from synthetic extracellular matrices

Kuen Yong Lee^{*†‡}, Martin C. Peters[‡], Kenneth W. Anderson^{*§}
& David J. Mooney^{*†‡}

Departments of ^{*}Biologic & Materials Sciences, [†]Chemical Engineering, [‡]Biomedical Engineering, and [§]Otolaryngology—Head & Neck Surgery, University of Michigan, Ann Arbor, Michigan 48109, USA

Polymeric matrices can be used to grow new tissues and organs^{1,2}, and the delivery of growth factors from these matrices is one method to regenerate tissues^{3,4}. A problem with engineering tissues that exist in a mechanically dynamic environment, such as bone, muscle and blood vessels^{5,6}, is that most drug delivery systems have been designed to operate under static conditions. We thought that polymeric matrices, which release growth factors in response to mechanical signals, might provide a new approach to guide tissue formation in mechanically stressed environments. Critical design features for this type of system include the ability to undergo repeated deformation, and a reversible binding of the protein growth factors to polymeric matrices to allow for responses to repeated stimuli. Here we report a model delivery system that can respond to mechanical signalling and upregulate the release of a growth factor to promote blood vessel formation. This approach may find a number of applications, including regeneration and engineering of new tissues and more general drug-delivery applications.

Natural extracellular matrices (ECMs) of tissues are regarded as depots for various growth factors, which are released to cells in the surrounding tissue to affect many physiological processes⁷. For example, release of vascular endothelial growth factor (VEGF) specifically enhances vascularization of tissues⁸. Synthetic and naturally derived polymers can be used similarly as depots and delivery vehicles of protein growth factors⁹. However, most tissues in the body are subjected to mechanical stimuli, and this form of signalling should be considered in the design of polymeric matrices

that release growth factors. We thought that growth factor delivery could be designed to respond to compressive stimulation of the matrix, owing to the increased pressure in the carrier resulting from matrix deformation. This mechanically controlled delivery of growth factors might allow the local environment of the delivery device to regulate the local concentration of the factor. This would be particularly advantageous when using growth factor delivery to engineer new vascular tissues or bone. We tested this hypothesis using hydrogel carriers, as hydrogels are frequently used in tissue engineering¹⁰ and are capable of repeated deformation following compressional loading.

To confirm that compressional loading can regulate drug release from polymeric matrices, we prepared alginate hydrogels containing trypan blue as a model drug molecule and subjected them to repeated compressive loading using a mechanical tester. The release rate and cumulative release of the drug increased during each incident of mechanical loading; however, the cumulative release at the end of each relaxation period was similar to that of control (non-compressed) hydrogels (see Supplementary Information). In this situation, hydrogels do not bind this drug, and the mechanical signalling does not contribute significantly to the release behaviour, owing to rapid depletion of the incorporated drug from the hydrogel.

We subsequently proposed that reversible binding of drugs to the hydrogel carrier, similar to the binding of growth factors to natural ECMs, would allow for a significant effect of cyclic mechanical stimulation on the factor release. VEGF was chosen as a model growth factor because of its use in many clinical indications and its reversible binding interaction with polysaccharides⁸. In the absence of mechanical stimulation, the release rate was fairly constant and the cumulative release of VEGF from the alginate hydrogels increased linearly with time. However, the release rate under mechanical stimulation increased to a value up to five times higher than that of control gels, and this increase could be regulated by the amplitude of the compression (Fig. 1). In this situation, free drug (not bound) would be released after each deformation, but the system would re-equilibrate during the subsequent relaxation by the disassociation of bound drug from the hydrogel. The cumulative release profile showed a stepwise increment with mechanical signalling, and the total amount of VEGF released from the hydrogels was up to two times

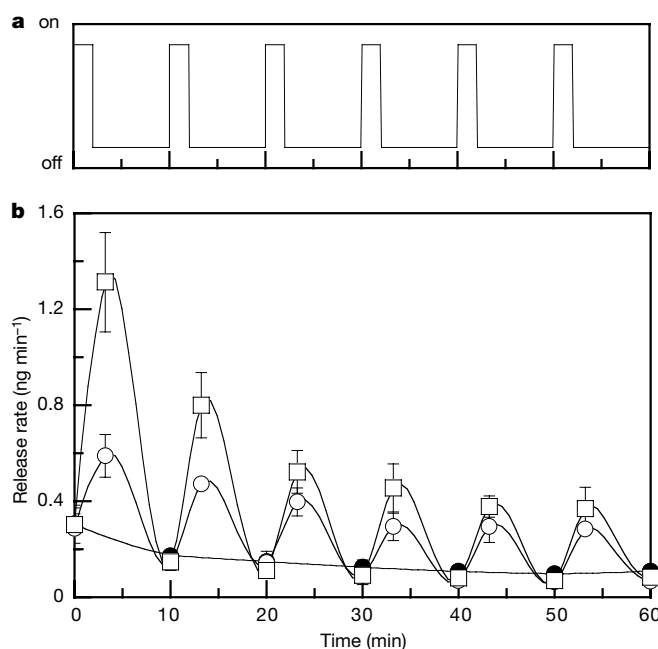


Figure 1 *In vitro* release profile of VEGF from alginate hydrogels under mechanical stimulation. **a**, Mechanical stimulation involving six cycles of compression for 2 min, followed by relaxation for 8 min. **b**, Release rate of VEGF from alginate hydrogels under

10% ($n = 4$; open circles) and 25% ($n = 4$; open squares) strain amplitude, and with no compression ($n = 4$; filled circles) as a control.

more than that of control conditions, depending on the magnitude of the mechanical input (see Supplementary Information).

Next, we tested the ability of mechanical stimulation to upregulate growth factor release *in vivo* and affect a local tissue response to show the efficacy and utility of this concept. We used VEGF as a model factor as it is a potent endothelial-cell-specific mitogen; it also enhances neovascularization, which accompanies many physiological and pathological processes¹¹, resulting in an increase of granulation tissue thickness and of the number of blood vessels. We subcutaneously implanted hydrogels loaded with VEGF into the dorsal region of severe combined immunodeficient (SCID) mice; we also implanted hydrogels without growth factors as a control. In the absence of growth factors, no remarkable vascularization surrounding the implant was observed with or without mechanical stimulation (Fig. 2a, b). The tissue sections surrounding hydrogels loaded with VEGF showed enhanced vascularization as expected (Fig. 2c, d). Notably, VEGF-loaded hydrogels subjected to cyclic mechanical stimulation showed a statistically significant increase in granulation layer thickness and in vascularization, as compared with non-stimulated VEGF-releasing gels. The thickness of granulation tissue surrounding the implants increased from 33 ± 8 to $74 \pm 9 \mu\text{m}$ (mean $\pm$ s.d. $P < 0.01$; Fig. 3a), and the number of blood vessels in the granulation tissue layer increased from 60 ± 10 to 91 ± 5 under mechanical stimulation ($P < 0.05$; Fig. 3b).

We investigated this approach further in non-obese diabetic (NOD) mice, as these diabetic animals are frequently used as an animal model for impaired wound healing in humans. We implanted alginate hydrogels containing VEGF into NOD mice after femoral artery ligation to discern whether an increase in collateral circulation could be achieved. Blood vessel formation was quantified after 14 d of implantation, and enhanced blood vessel formation was observed in mechanically stimulated implant sites (Fig. 4a, b). The density of blood vessels increased from 161 ± 32 to $260 \pm 65 \text{ mm}^{-2}$ under cyclic mechanical stimulation ($P < 0.05$; Fig. 4c). Hydrogels without VEGF were also used as a control, and no significant blood vessel formation was observed.

Our findings have a number of potential applications for tissue engineering and drug delivery, and also indicate a new general mechanism by which tissues in the body may respond to mechanical signals. It has been proposed that stress-responsive mechanisms include alterations in cellular structure or metabolism¹², and pre-

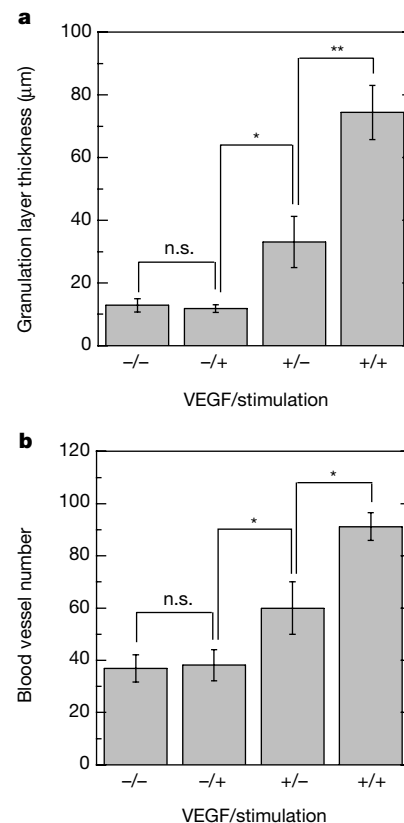


Figure 3 Quantitative analysis of granulation tissue formed in SCID mice. Shown are thickness of granulation tissue (**a**) and the number of blood vessels (**b**) at experimental conditions of no VEGF/no mechanical stimulation (-/-), no VEGF/mechanical stimulation (-/+), VEGF incorporated in hydrogel/no mechanical stimulation (+/-) and VEGF incorporated in hydrogel/mechanical stimulation (+/+) ($n = 4$). n.s., no statistical difference. Asterisk indicates statistical significance, $P < 0.05$. Double asterisk indicates statistical significance, $P < 0.01$.

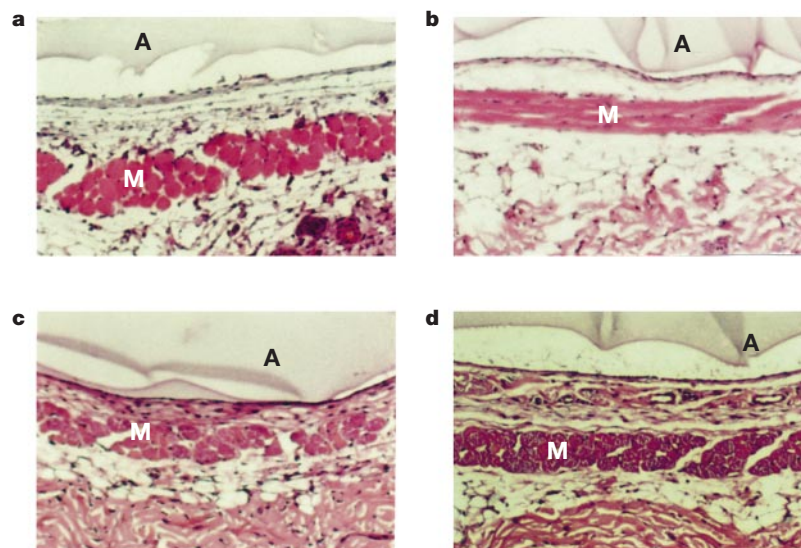


Figure 2 *In vivo* response to VEGF released from alginate hydrogels under mechanical stimulation. The gels were implanted into the dorsal region of SCID mice for 14 d. Photomicrographs of representative tissue section from each experimental condition: no VEGF/no mechanical stimulation (**a**); no VEGF/mechanical stimulation (**b**); VEGF

incorporated in hydrogel/no mechanical stimulation (**c**); and VEGF incorporated in hydrogel/mechanical stimulation (**d**). Original magnification $\times 100$. A, alginate hydrogel; M, muscle layer.

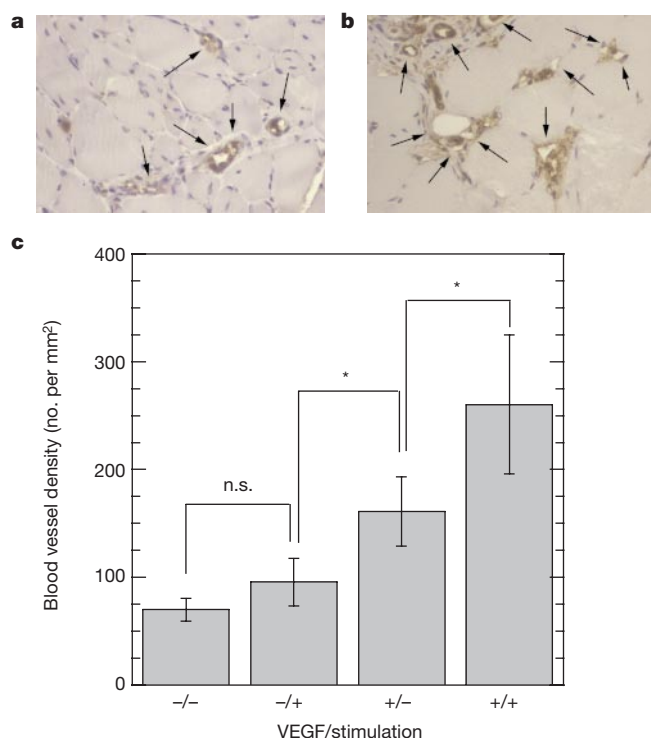


Figure 4 *In vivo* response to VEGF-loaded hydrogels implanted into femoral artery ligation site of NOD mice. Shown are photomicrographs of representative tissue section: **a**, VEGF incorporated in hydrogel/no mechanical stimulation (+/-); **b**, VEGF incorporated in hydrogel/mechanical stimulation (+/+). **c**, Blood vessel density (number per mm²) at each experimental condition ($n = 4$). Control gels without VEGF under static conditions (-/-) and under mechanical stimulation (-/+) were also implanted. Original magnification $\times 400$. Arrows indicate CD31-stained blood vessels. n.s., no statistical difference. Asterisk indicates statistical significance, $P < 0.05$.

vious studies that focused on the role of growth factors have detailed changes in transcription and/or translation of the factors¹³. It is possible, perhaps even likely, that the application of mechanical stress to tissues will lead to an increased release of growth factors through the mechanism that we describe. This may provide an important and previously unrecognized means by which stressed tissues communicate with surrounding tissues, and adapt to their environment. As such, the systems that we have described here may be simply mimicking a critical function of natural ECMs. It may also be possible to develop delivery systems of small drug molecules using this concept. There have been many attempts to develop drug-delivery systems in response to external stimuli such as temperature, pH, ultrasound, and electric or magnetic field^{14–19}. However, little effort has been devoted to mechanical stimuli-mediated drug-delivery systems, relative to the significance of these signals in physiological situations. This approach could allow a high level of control over the delivery dose and rate, and might even be incorporated into feedback systems to allow highly precise delivery profiles. □

Methods

In vitro release

We prepared alginate hydrogels (10 g, Pronova) containing 63 mg CaSO₄, with either trypan blue (4 mg, Sigma) or VEGF (10 µg, Intergen). We included ¹²⁵I-labelled VEGF (Biomedical Technologies) as a tracer in certain experiments. Gels were cut into disks (12.7 mm diameter and 2 mm thick), and pre-swollen in DMEM overnight. The gel disks were put into custom-made sample holders (13 mm diameter and 30 mm length) with porous stainless-steel supports on the top and bottom of the hydrogels to allow the unimpeded release of the drug from the hydrogel. We compressed the hydrogels with a mechanical tester (MTS, France), and applied 3–6 cycles of compressive loading (10 or 25% strain; 2-min strain and 8-min relaxation). The amount of released trypan blue was spectrophotometrically determined at 588 nm. We measured the release of radiolabelled VEGF by a gamma counter (Packard), and converted it to absolute protein amounts using

the known specific activity of ¹²⁵I-labelled VEGF.

Animal models

Alginate hydrogels (0.5 g) containing 10 µg VEGF and 3.2 mg CaSO₄ were cut into disks (4.8 mm diameter and 2 mm thick) and pre-swollen in DMEM overnight. We anaesthetized animals with an intramuscular injection of 87 mg per ml ketamine and 2.6 mg per ml xylazine. The gel disks were subcutaneously implanted into the dorsal region of 7–9-week-old SCID mice in the first model, and a 24-h recovery period was allowed. Disks were subsequently stimulated daily for 7 d using 50% strain amplitude at three cycles of compression for 1 min, followed by relaxation for 1 min using a custom-made stimulator.

In the second model, the femoral artery in the right hind limb of 7–9-week-old NOD mice was ligated and the exposed arterial ends were tied off with nylon sutures. We then placed a single hydrogel disk directly on the ligation site of each animal and closed the skin with sutures. Twenty-four hours after surgery, we mechanically stimulated the implantation site daily for 7 d, at six cycles of compression for 30 s, followed by relaxation for 90 s. The treatment of experimental animals was in accordance with University of Michigan animal care guidelines, and we observed all NIH animal handling procedures.

After 14 d, the mice in both models were killed and tissues directly surrounding each implant were taken. The tissues were fixed in zinc-formalin solution at 4°C overnight, dehydrated through graded ethanol, embedded in paraffin and cut into 4-µm sections. To identify blood vessels, the tissue sections from SCID mice were stained with haematoxylin and eosin, and those from NOD mice were immunostained with antibodies raised against mouse CD31 (Pharmingen) using a terminator blocking solution, universal link secondary antibody, streptavidin/horseradish peroxidase (HRP) (Biocare Medical) and 3,3'-diaminobenzidine (Zymed) as described.²⁰

We quantified the number of blood vessels and the granulation layer thickness by capturing digital images of tissue sections and analysing them with NIH Image software. The thickness of the newly formed granulation tissue layer was determined between hydrogels and muscle layers. Granulation layer thickness and the number of blood vessels in SCID mice were quantified at $\times 100$ and $\times 400$ magnifications, respectively, and normalized by the interfacial length between the hydrogel and the granulation tissue. We counted the density of blood vessels formed in NOD mice at $\times 400$ magnification. We carried out statistical analysis using Instat software.

Received 11 May; accepted 25 September 2000.

- Langer, R. & Vacanti, J. P. Tissue engineering. *Science* **260**, 920–926 (1993).
- Putnam, A. J. & Mooney, D. J. Tissue engineering using synthetic extracellular matrices. *Nature Med.* **2**, 824–826 (1996).
- Shea, L. D., Smiley, E., Bonadio, J. & Mooney, D. J. DNA delivery from polymer matrices for tissue engineering. *Nature Biotechnol.* **17**, 551–554 (1999).
- Ripamonti, U. & Reddi, A. H. Tissue engineering, morphogenesis, and regeneration of the periodontal tissues by bone morphogenetic proteins. *Crit. Rev. Oral Biol. Med.* **8**, 154–163 (1997).
- Kim, B.-S., Nikolovski, J., Bonadio, J. & Mooney, D. J. Cyclic mechanical strain regulates the development of engineered smooth muscle tissue. *Nature Biotechnol.* **17**, 979–983 (1999).
- Niklason, L. E. *et al.* Functional arteries grown *in vitro*. *Science* **284**, 489–493 (1999).
- Vlodavsky, I. *et al.* Extracellular sequestration and release of fibroblast growth factor: a regulatory mechanism? *Trends Biochem. Sci.* **16**, 268–271 (1991).
- Neufeld, G., Cohen, T., Gengrinovitch, S. & Poltorak, Z. Vascular endothelial growth factor (VEGF) and its receptors. *FASEB J.* **13**, 9–22 (1999).
- Baldwin, S. P. & Saltzman, W. M. Materials for protein delivery in tissue engineering. *Adv. Drug Delivery Rev.* **33**, 71–86 (1998).
- Jen, A. C., Wake, M. C. & Mikos, A. G. Hydrogels for cell immobilization. *Biotechnol. Bioeng.* **50**, 357–364 (1996).
- Plate, K. H., Breiser, G., Weich, H. A. & Risau, W. Vascular endothelial growth factor is a potential tumor angiogenesis factor *in vivo*. *Nature* **359**, 845–848 (1992).
- Chicurel, M. E., Chen, C. S. & Ingber, D. E. Cellular control lies in the balance of forces. *Curr. Opin. Cell Biol.* **10**, 232–239 (1998).
- Williams, B. Mechanical influences on vascular smooth muscle cell function. *J. Hypertension* **16**, 1921–1929 (1998).
- Langer, R. Drug delivery and targeting. *Nature* **392** (suppl.), 5–10 (1998).
- Wang, C., Stewart, R. J. & Kopecek, J. Hybrid hydrogels assembled from synthetic polymers and coiled-coil protein domains. *Nature* **397**, 417–420 (1999).
- Chen, G. & Hoffman, A. S. Graft copolymers that exhibit temperature-induced phase transition over a wide range of pH. *Nature* **373**, 49–52 (1995).
- Mitragori, S., Blankschtein, D. & Langer, R. Ultrasound-mediated transdermal protein delivery. *Science* **269**, 850–853 (1995).
- Kwon, I. C., Bae, Y. H. & Kim, S. W. Electrically erodible polymer gel for controlled release of drugs. *Nature* **354**, 291–293 (1991).
- Edelman, E., Brown, L. & Langer, R. *In vitro* and *in vivo* kinetics of regulated drug release from polymer matrices by oscillating magnetic fields. *J. Biomed. Mater. Res.* **21**, 339–353 (1987).
- Nör, J. E., Christensen, J., Mooney, D. J. & Polverini, P. J. Vascular endothelial growth factor (VEGF)-mediated angiogenesis is associated with enhanced endothelial cell survival and induction of BCL-2 expression. *Am. J. Pathol.* **152**, 375–384 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank the National Institutes of Health for financial support of this research. M.C.P. acknowledges the Whitaker Foundation for a graduate fellowship.

Correspondence and requests for materials should be addressed to D.J.M. (e-mail: mooneyd@umich.edu).

A role for *Saccharomyces cerevisiae* histone H2A in DNA repair

Jessica A. Downs*, Noel F. Lowndes† & Stephen P. Jackson*

* The Wellcome Trust and Cancer Research Campaign, Institute of Cancer and Developmental Biology, and Department of Zoology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QR, UK

† ICRF Clare Hall Laboratories, CDC Laboratory, Blanche Lane, South Mimms, Potters Bar, EN6 3LD, UK

Histone proteins associate with and compact eukaryotic nuclear DNA to form chromatin. The basic unit of chromatin is the nucleosome, which is made up of 146 base pairs of DNA wrapped around two of each of four core histones¹, H2A, H2B, H3 and H4. Chromatin structure and its regulation are important in transcription and DNA replication^{2–4}. We therefore thought that DNA-damage signalling and repair components might also modulate chromatin structure. Here we have characterized a conserved motif in the carboxy terminus of the core histone H2A from *Saccharomyces cerevisiae* that contains a consensus phosphorylation site for phosphatidylinositol-3-OH kinase related kinases (PIKKs). This motif is important for survival in the presence of agents that generate DNA double-strand breaks, and the phosphorylation of this motif in response to DNA damage is dependent on the PIKK family member Mec1. The motif is not necessary for Mec1-dependent cell-cycle or transcriptional responses to DNA damage, but is required for efficient DNA double-strand break repair by non-homologous end joining. In addition, the motif has a role in determining higher order chromatin structure. Thus, phosphorylation of a core histone in response to DNA damage may cause an alteration of chromatin structure that facilitates DNA repair.

The PIKK family of kinases, which includes the human proteins DNA-dependent protein kinase (DNA-PK), ataxia telangiectasia-mutated (ATM) and AT-related (ATR), and the *S. cerevisiae* ATR/ATM homologues, Mec1 and Tel1, is important for DNA-damage signalling in eukaryotes^{5,6}. Although these kinases are central in the response to DNA damage, their downstream targets have not been completely identified. Ser-Gln-Glu (SQE) is a consensus site that serves as an efficient substrate for phosphorylation *in vitro* by DNA-PK and ATM^{7–10}. An SQE motif exists at the extreme C terminus of several histone H2A proteins (Fig. 1a). The conservation of this motif suggested that its phosphorylation by PIKKs might act in DNA repair and/or DNA-damage signalling. Consistent with this idea, phosphorylation of this motif is observed in response to ionizing radiation in several organisms¹¹.

To determine whether the SQE motif of yeast histone H2A is important for DNA-damage responses, we used a system in which both genomic H2A and H2B copies have been deleted; the strain is kept alive by the presence of one set of H2A/H2B genes on a plasmid (strain FY406; ref. 12). In this background, we constructed a strain in which Ser 129 of H2A was changed to a stop codon (*hta1*-S129*), thereby removing the four C-terminal amino-acid residues, SQEL. The *hta1*-S129* strain is hypersensitive to the DNA-damage-inducing agents phleomycin and methyl methane-sulphonate (MMS; Fig. 1b, c), indicating that the four C-terminal amino-acid residues of yeast H2A may be important for the signalling and/or repair of DNA damage.

To assess the contributions of each residue within the C-terminal motif, we created a series of H2A mutant strains. When either Ser 129 or Gln 130 is replaced with alanine (S129A and Q130A, respectively), the resulting strain is impaired for growth on MMS-containing plates to a similar degree as *hta1*-S129* (Fig. 1c). Replacement of Glu 131 with alanine (E131A) results in slight

hypersensitization; changing Leu 132 to alanine (L132A) yields a phenotype intermediate between the S129*/S129A/Q130A mutations and the E131A mutation. These phenotypes correlate with the relative importance of these residues in defining an optimal PIKK recognition motif, and suggest that MMS hypersensitivity may reflect loss of PIKK recognition. We mutated Ser 129 to either threonine or glutamic acid to maintain phosphorylation ability or to mimic constitutive phosphorylation, respectively. Strains carrying either of these mutations survive in the presence of MMS almost as well as the wild-type strain (Fig. 1c), consistent with phosphorylation of this residue being important for the resistance phenotype. Notably, none of the *hta1*-mutant strains is hypersensitive to ultraviolet irradiation or ethyl methane-sulphonate (EMS; Fig. 1d; and data not shown). This suggests that the C terminus of histone H2A may be specifically important for survival of DNA double-strand breaks (DSBs), as MMS and phleomycin, to which *hta1*-S129* strain is hypersensitive, cause DSBs in DNA^{13,14}. There are no consistent differences between the levels of mutant histone H2As (Fig. 1e), suggesting that the phenotypes observed are not due to protein instability.

To study the phosphorylation status of histone H2A, we generated antibodies specific to phosphorylated Ser 129. These were used

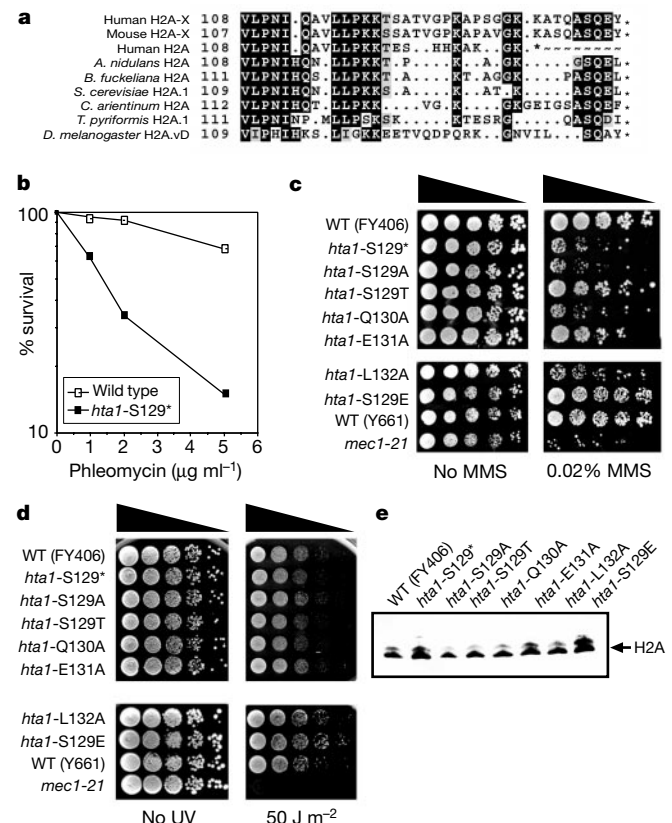


Figure 1 The four C-terminal amino-acid residues (SQEL) of histone H2A are important for survival after DNA damage. **a**, Alignment of histone H2A C termini from *Homo sapiens* (human H2A and H2A-X), *Mus musculus* (mouse H2A-X), *Aspergillus nidulans* (H2A), *Botrytis fuckeliana* (H2A), *Saccharomyces cerevisiae* (H2A.1), *Cicer arietinum* (H2A), *Tetrahymena pyriformis* (H2A.1) and *Drosophila melanogaster* (H2A.vD). The amino-acid position in each protein is shown on the left. Sequences shown are C-terminal to the conserved core histone fold; stop codons are indicated by asterisks. **b**, Survival of wild-type (FY406) and *hta1*-S129* mutant yeast on phleomycin-containing plates. **c**, Serial dilutions of *hta1*-mutant yeast strains and their isogenic wild-type control (FY406) grown with or without 0.02% MMS. **d**, Serial dilutions of *hta1*-mutant yeast strains and their control (FY406) plated on medium; one set was exposed to ultraviolet irradiation. **e**, Western blot analysis of yeast lysates prepared from indicated strains and analysed using polyclonal antiserum against recombinant human histone H2A-X.

to probe western blots of yeast protein extracts. Although similar amounts of histone H2A are detected in extracts from a wild-type strain grown with or without MMS, the phospho-specific antiserum only recognizes histone H2A after growth in MMS or phleomycin (Fig. 2a; and data not shown). These results agree with previous observations¹¹ and show that yeast histone H2A is phosphorylated on Ser 129 *in vivo* in response to DNA damage.

Because histone H2A Ser 129 lies within a PIKK recognition motif, we tested whether the two known yeast DNA-damage-responsive PIKKs, Mec1 and Tel1, are required for Ser 129 phosphorylation. We assayed strains bearing mutations in *MEC1* (*mec1-21*), *TEL1* (*tel1::HIS3*) and an isogenic control¹⁵ for phosphorylated H2A in the presence or absence of MMS. Although signal was detected in wild-type and *tel1* mutant yeast, no phosphorylation of H2A was observed in the *mec1* or *mec1/tel1* mutant cells (Fig. 2b). On very long exposures, a low level of signal in the

mec1, but not *mec1/tel1*, mutant cells was detected (data not shown), indicating some residual phosphorylation by Tel1 in *mec1* mutant cells. In contrast, DNA-damage-dependent phosphorylation of H2A was not lost in strains with mutations in other components of the Mec1-dependent DNA-damage signalling pathway (Fig. 2c). These data suggest that H2A may be a direct target for Mec1. Consistent with this, immunoprecipitated Mec1 from yeast cell extracts can phosphorylate an H2A C-terminal peptide *in vitro* (see Supplementary Information). Recent work in mammalian cells has shown that the DNA-damage-dependent phosphorylation of the human H2AX SQE motif is greatly impaired by the PIKK family inhibitor wortmannin¹⁶, suggesting that this conserved histone motif is a target for the PIKKs throughout eukaryotic evolution.

A set of *S. cerevisiae* genes is transcriptionally activated by DNA damage, and this response is dependent on Mec1 (ref. 17). To see whether Mec1-dependent phosphorylation of histone H2A is important for activation of these genes, we analysed the expression of two DNA-damage response genes, *RNR2* and *RNR4* (ref. 18). The DNA-damage-dependent transcriptional activation of these genes

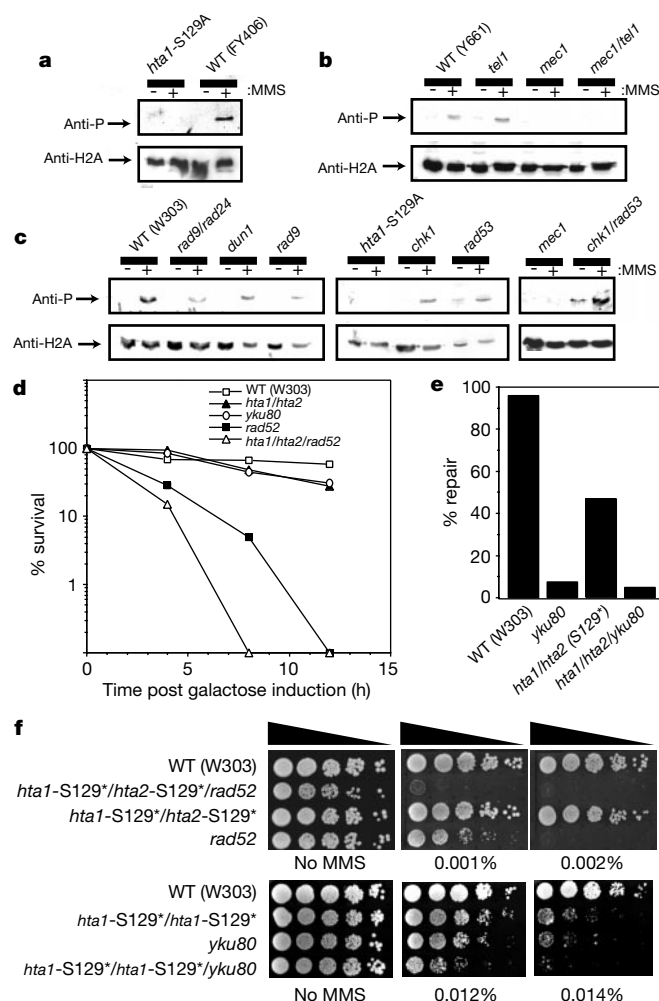


Figure 2 Serine 129 of histone H2A is phosphorylated in a Mec1-dependent manner and is important for NHEJ. **a**, Western blot of wild-type (FY406) and *hta1-S129A* lysates made from cultures grown with (+) or without (–) 0.05% MMS. Top panel was analysed with antiserum against the H2A C-terminal phosphopeptide; bottom panel was analysed with polyclonal antiserum against human histone H2A-X. **b**, **c**, Western blot of lysates prepared from the indicated strains, analysed as in **a**. The H2A phosphorylation in the absence of exogenously added damage in *rad53* mutant strains may be due to increased levels of unrepaired DNA damage in these strains. **d**, Survival of strains containing a plasmid with a galactose-inducible HO endonuclease expressed relative to survival at time zero. **e**, The plasmid NHEJ assay was performed on the indicated strains. Repair is expressed as the percentage of colony formation for transformation with linear versus uncut plasmid. **f**, Serial dilutions of the indicated yeast strains and their isogenic wild-type control (W303) plated on medium with the indicated concentration of MMS.

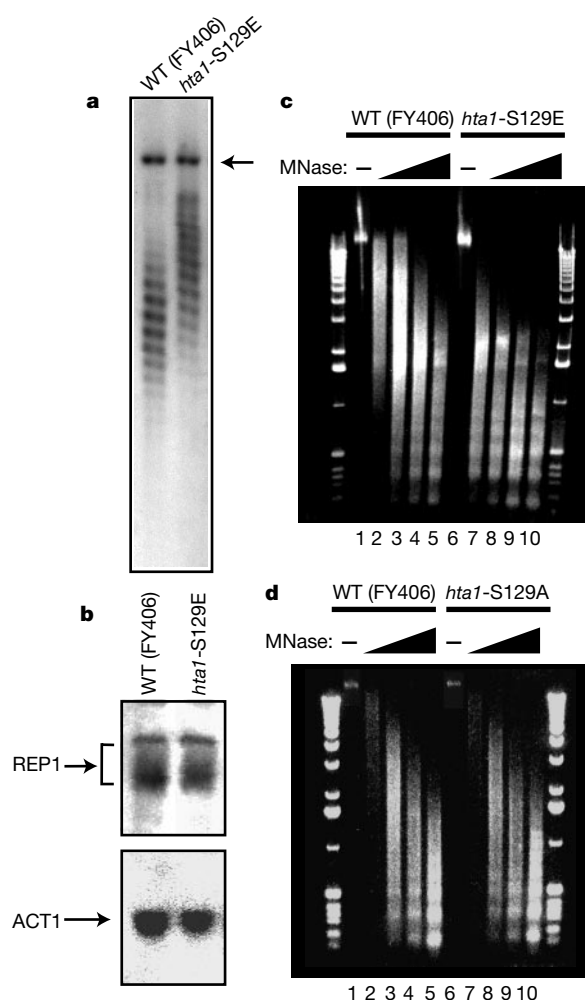


Figure 3 Substitution of H2A Ser 129 with glutamic acid changes global chromatin structure. **a**, DNA as isolated from wild-type or *hta1-S129E* strains and electrophoresed on a chloroquine-containing gel. 2μ plasmid superhelical density was analysed by Southern blotting and hybridization with a radiolabelled probe. Open-circular plasmid is indicated with an arrow. **b**, Northern blot analysis of the 2μ-encoded *REP1* transcripts and the genomic *ACT1* transcript from wild-type and *hta1-S129E*-containing strains. **c**, Nuclei were isolated from wild-type or *hta1-S129E* cells and incubated with no MNase (lanes 1 and 6) or 2.86 units of MNase for increasing periods of time. The DNA was isolated, electrophoresed, and visualized by staining with ethidium bromide. **d**, Nuclei isolated from wild-type and *hta1-S129A* cells were analysed as in **c**.

in wild-type and the *hta1*-S129* mutant cells is similar (see Supplementary Information), suggesting that the sensitivity of the H2A mutant strain to DNA damage is not caused by a transcriptional defect. Mec1 is also important for the DNA-damage cell-cycle checkpoints^{6,19}. We therefore compared the abilities of wild type (W303) and *hta1*-S129*/*hta2*-S129* mutant yeast (JDY22; see Methods) to arrest cell-cycle progression in response to MMS, but found no detectable differences between the two strains (see Supplementary Information). It therefore appears that strains lacking the C-terminal PIKK motif of histone H2A are not overtly checkpoint defective.

Another possibility is that the histone H2A C terminus facilitates survival by directly affecting the repair of DNA lesions. DNA DSB repair in eukaryotes is mediated primarily by two pathways²⁰: homologous recombination, which is dependent on *RAD52*; and non-homologous end joining (NHEJ), which requires *YKU70* and *YKU80*. To determine whether either pathway is affected by the loss of the H2A C terminus, we assayed the ability of the *hta1*-S129*/*hta2*-S129* strain (JDY22) to grow in the presence of the homothallic switching (HO) endonuclease by using a galactose-inducible construct. Strains deficient in homologous recombination, but not in NHEJ, are sensitive to galactose-containing medium when carrying this plasmid²¹. Survival of the *rad52* mutant strain was extremely impaired when compared with its wild-type control (Fig. 2d). In contrast, and similar to a *yku80* mutant strain, the H2A mutant strain was only slightly hypersensitive to HO expression (Fig. 2d). This indicates that homologous recombination is not substantially impaired in the absence of the histone H2A C-terminal SQEL motif. Furthermore, the *hta1*-S129*/*hta2*-S129*/*rad52* mutant strain was more sensitive to HO endonuclease expression than the *rad52* strain alone (Fig. 2d), suggesting that histone H2A can facilitate DSB repair in a manner that is distinct from the *RAD52*-dependent homologous recombination pathway.

We next examined the role of histone H2A in DNA NHEJ by using a plasmid repair assay, in which repair of a linear plasmid that cannot be mediated efficiently by homologous recombination is measured²². Although not to the same degree as a *yku80* mutant strain, we found that the *hta1*-S129*/*hta2*-S129* mutant strain is deficient in NHEJ as compared with its wild-type control (Fig. 2e). It has been shown that the Sir proteins function in NHEJ, in part, by maintaining haploid-specific gene regulation through silencing at the *HMR* and *HML* loci^{23,24}. To see whether histone H2A similarly influences NHEJ, we introduced the *hta1*-S129*/*hta2*-S129* mutations into a strain lacking *HMR* and *HML* (JKM115; ref. 24). In this strain, the levels of plasmid repair relative to the isogenic wild-type strain were reduced to roughly the same level as in the strain containing wild-type *HML* and *HMR* (see Supplementary Information), indicating that the impaired levels of NHEJ in the H2A-mutant strain are not caused by a misregulation of haploid-specific genes. We also found that the levels of NHEJ in the *hta1*-S129*/*hta2*-S129*/*ku80* mutant strain are lower than in either the *hta1*-S129*/*hta2*-S129* strain or the *ku80* mutant strain (Fig. 2e). It is therefore possible that the H2A C terminus could also facilitate a third DSB repair pathway, work in homologous recombination in a manner that is not detectable in the galactose–HO assay, or facilitate NHEJ in a manner that is not epistatic with Ku80. Consistent with this, the MMS sensitivity of histone H2A S129* is not epistatic with *RAD52* or *YKU80* (Fig. 2f).

One possible mechanism by which phosphorylation of a histone might facilitate DNA repair is by altering chromatin structure surrounding the DNA lesion in a manner such that subsequent repair events are facilitated. To investigate this possibility, we took advantage of the fact that if Ser 129 is replaced by glutamic acid, yeast survive DNA damage to a similar extent as wild type (Fig. 1c). This indicates that Glu 129 can replace the function(s) normally provided by phosphorylation of Ser 129 that are important for viability in the presence of DNA damage. Therefore, we compared

the chromatin structure of the *hta1*-S129E-containing strain with wild-type yeast bearing an unmodified Ser 129. Because wrapping of DNA around a histone octamer introduces a negative supercoil into a closed circular plasmid²⁵, the superhelical density of the endogenous 2 μ plasmid can provide information about changes in chromatin structure²⁶. We found that the profile of topoisomers from *hta1*-S129E-containing cells is consistently more relaxed relative to the profile seen in wild-type cells (Fig. 3a). When we examined the 2 μ -encoded *REP1* transcript levels in wild-type and *hta1*-S129E mutant strains, we detected no significant differences in *REP1* transcript levels relative to the genomic *ACT1* control, suggesting that 2 μ transcription levels are not grossly altered in the *hta1*-S129E mutant strain (Fig. 3b). However, other aspects of 2 μ biology might be altered by the *hta1*-S129E mutation and subsequently affect topoisomer distribution.

Therefore, to examine further the potential differences in chromatin structure between these strains, we analysed genomic chromatin by micrococcal nuclease (MNase) digestion of nuclei isolated from wild-type and *hta1*-S129E cells. The digestion of higher molecular weight DNA occurred more rapidly in *hta1*-S129E-containing nuclei, and correspondingly, the lower molecular weight nucleosomal repeats appeared more slowly in wild-type nuclei (Fig. 3c, compare lanes 2 and 7). In contrast, no significant differences were detected between wild-type and *hta1*-S129A mutant strains (Fig. 3d), suggesting that the effect on chromatin structure is specific to a mutation that mimics phosphorylation. Notably, no differences were observed in either the pattern of protection or the lengths of individual nucleosomal repeats in the *hta1*-S129E-containing strain, suggesting that the histone octamers are not grossly altered in their mobility or interaction with DNA. Instead, the regions between individual nucleosomes appear to be more susceptible to degradation by MNase. This is consistent with the H2A C terminus being at the region of the nucleosome where DNA enters and exits^{1,27}, and suggests that higher order chromatin structure is influenced by the C terminus of histone H2A. Replacement of Ser 129 with a residue that mimics phosphorylation decreases chromatin compaction, suggesting that Mec1-dependent modification of Ser 129 may facilitate NHEJ by decondensing chromatin and thus allowing the repair factors better access to the DNA. □

Methods

Phleomycin sensitivity curves

Yeast strains were grown overnight at 30 °C in non-selective medium, diluted 1:50, and grown for 3 h in non-selective medium. Cultures were diluted to equivalent densities and plated onto media containing the indicated amounts of phleomycin. After incubation for 3 days, colonies were counted and displayed as a percentage relative to plates containing no phleomycin.

DNA-damaging agent assays

Strains were grown overnight at 30 °C in non-selective medium, diluted to a density corresponding to an absorbance (A) at 600 nm of 0.2, and grown for 3 h in non-selective medium. Cultures were diluted to equivalent densities, and fivefold serial dilutions were plated onto medium containing the indicated DNA damaging agent.

Western blot analysis

Extracts were prepared from mid-log cultures exposed to no MMS or 0.05% MMS for 2 h before lysis. Lysates were electrophoresed on 18% SDS polyacrylamide gels and transferred to nitrocellulose membranes, which were incubated with polyclonal antiserum directed against recombinant human H2A-X followed by detection using horseradish peroxidase (HRP)-conjugated anti-rabbit secondary antibody (Pierce) and visualization with enhanced chemiluminescence (Amersham). To detect phosphorylated H2A, polyclonal antiserum directed against a C-terminal 13-residue phosphopeptide was reverse purified on an unphosphorylated peptide column. The flow-through from this column was used in western blot analysis.

Galactose–HO sensitivity assay

Cells containing the galactose-inducible HO endonuclease were grown to mid-log, washed once with water and resuspended in selective medium containing either glucose or galactose as the carbon source. Samples were taken at various time points, the cell densities

measured by A_{600} , and cells plated on non-selective medium. The survival rate was calculated as the proportion of colony-forming units to total cells and was expressed relative to survival before galactose induction.

Plasmid repair assay

Cells were transformed with 1 μ g of pRS414 that was either mock digested or digested to completion with *Eco*RI. Transformations were plated onto selective medium using an automated plater (DW Scientific), and colonies were counted after 3 d at 30 °C. Repair is expressed as a ratio of colony formation from linear/circular plasmid transformations.

Yeast strains

Strains and their sources are as follows: Y662 (*tel1::HIS3*), Y663 (*mec1-21*) and Y664 (*mec1-21/tel1::HIS3*) are in the Y661 wild-type background (S. Elledge). *rad9::HIS3*, *rad9::HIS3/rad24::URA3*, *dun1::HIS3* and *rad53::HIS3* are in W303 (N. Lowndes). *chk1::HIS3* and *chk1::URA3/rad53::HIS3* are in W303 (J. Rouse). FY406-based strains (*hta1-htb1::LEU2*, *hta2-htb2::TRP1*, (pAB6); ref. 12) were made by site-directed mutagenesis of the wild-type *HTA1-HTA2* sequence that was cloned from pAB6 into pRS414. To generate strain JDY22, which has the Ser 129 to stop 129 mutation (S129*) introduced into both genomic copies of the H2A genes, a disruption construct containing the *hta1*-S129* mutation and a *hisG-URA3* cassette was transformed into W303 α . After selection on 5-FOA and confirmation of integration of the *hta1*-S129* mutation by polymerase chain reaction (PCR) and restriction enzyme analysis, this strain (JDY21) was transformed with a disruption construct containing the *hta2*-S129* mutation and a *hisG-URA3* cassette. Positive colonies were selected on 5-FOA and integration of the mutation was confirmed by PCR and restriction enzyme analysis. Disruption constructs for *YKU80* and *RAD52* were transformed into W303 α or JDY22 and confirmed by PCR to create strains JDY20 (*yku80::URA3*), JDY24 (*hta1*-S129*/*hta2*-S129*/*yku80::URA3*), JDY1 (*rad52::TRP1*) and JDY33 (*hta1*-S129*/*hta2*-S129*/*rad52::TRP1*).

2 μ topology assay

Superhelical density of the 2 μ plasmid was determined essentially as described²⁶. Briefly, cells were grown to mid-log phase and washed in toluene buffer (15 mM Tris-HCl, pH 8.0, 5 mM EDTA, pH 8.0, 1.5% toluene). Spheroplasts were prepared by incubation with Zymolyase for 15 min at 37 °C, and lysed by incubation with 1% SDS. DNA was isolated, electrophoresed on a 0.75% gel in 1 $\times$ TB buffer and 0.6 μ g ml⁻¹ chloroquine, and transferred to nylon membrane (Amersham). PCR product was amplified using primers against a 2-kb region of the 2 μ plasmid and was labelled by random priming (Promega). Hybridization was visualized by phosphorimager analysis.

MNase digestion

Nuclei were isolated from the indicated strains as described²⁶. These were normalized and resuspended in 100 μ l MNase buffer (1 M sorbitol, 20 mM PIPES, pH 6.5, 5 mM CaCl₂, 150 mM NaCl), 2.86 U MNase was added, and the samples were incubated for 0, 1, 2, 5 or 10 min. Reactions were terminated by addition of 10 μ l MNase stop buffer (0.5 M Tris-HCl, pH 8.0, 250 mM EDTA, 5% SDS) and 10 μ g Proteinase K. The samples were incubated for 2 h at 50 °C, then phenol/chloroform extracted twice, chloroform extracted, and ethanol precipitated in the presence of glycogen. Samples were electrophoresed on a 1% agarose gel and visualized by staining with ethidium bromide.

Northern blot analysis

RNA was prepared from wild-type and *hta1*-S129E-containing cells using RNeasy (Promega), electrophoresed and transferred to nylon membrane (Amersham). PCR products corresponding to the *REP1* or *ACT1* open reading frames were labelled by random priming (Promega). Hybridization was visualized by phosphorimager analysis.

Received 26 June; accepted 12 October 2000.

- Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
- Wolffe, A. P. & Hayes, J. J. Chromatin disruption and modification. *Nucleic Acids Res.* **27**, 711–720 (1999).
- Kornberg, R. D. & Lorch, Y. Chromatin-modifying and -remodeling complexes. *Curr. Opin. Genet. Dev.* **9**, 148–151 (1999).
- Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349–352 (1997).
- Smith, G. C. M. & Jackson, S. P. The DNA-dependent protein kinase. *Genes Dev.* **13**, 916–934 (1999).
- Weinert, T. DNA damage and checkpoint pathways: molecular anatomy and interactions with repair. *Cell* **94**, 555–558 (1998).
- O'Neill, T. et al. Utilization of oriented peptide libraries to identify substrate motifs selected by ATM. *J. Biol. Chem.* **275**, 22719–22727 (2000).
- Kim, S.-T., Lim, D.-S., Canman, C. E. & Kastan, M. B. Substrate specificities and identification of putative substrates of ATM kinase family members. *J. Biol. Chem.* **274**, 37538–37543 (1999).
- Anderson, C. W. & Lees-Miller, S. P. The nuclear serine/threonine protein kinase DNA-PK. *Crit. Rev. Eukaryotic Gene Exp.* **2**, 283–314 (1992).
- Bannister, A. J., Gottlieb, T. M., Kouzarides, T. & Jackson, S. P. c-Jun is phosphorylated by the DNA-dependent protein kinase in vitro: definition of the minimal kinase recognition motif. *Nucleic Acids Res.* **21**, 1289–1295 (1993).
- Rogakou, E. P., Boon, C., Redon, C. & Bonner, W. M. Megabase chromatin domains involved in DNA double-strand breaks in vivo. *J. Cell Biol.* **146**, 905–915 (1999).
- Hirschhorn, J. N., Bortvin, A. L., Ricupero-Hovasse, S. L. & Winston, F. A new class of histone H2A mutations in *Saccharomyces cerevisiae* causes specific transcriptional defects in vivo. *Mol. Cell Biol.* **15**, 1999–2009 (1995).

- Schwartz, J. L. Monofunctional alkylating agent-induced S-phase-dependent DNA damage. *Mutat. Res.* **216**, 111–118 (1989).
- Povirk, L. F. DNA damage and mutagenesis by radiomimetic DNA-cleaving agents: bleomycin, neocarzinostatin and other enediynes. *Mutat. Res.* **355**, 71–89 (1996).
- Sanchez, Y. et al. Regulation of *RAD53* by the *ATM*-like kinases *MEC1* and *TEL1* in yeast cell cycle checkpoint pathways. *Science* **271**, 357–360 (1996).
- Paull, T. T. et al. A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr. Biol.* **10**, 886–895 (2000).
- de la Torre-Ruiz, M. A., Green, C. M. & Lowndes, N. F. *RAD9* and *RAD24* define two additive, interacting branches of the DNA damage checkpoint pathway in budding yeast normally required for *Rad53* modification and activation. *EMBO J.* **17**, 2687–2698 (1998).
- Elledge, S. J., Zhou, Z., Allen, J. B. & Navas, T. A. DNA damage and cell cycle regulation of ribonucleotide reductase. *BioEssays* **15**, 333–339 (1993).
- Longhese, M. P., Foiani, M., Muzi-Falconi, M., Lucchini, G. & Plevani, P. DNA damage checkpoint in budding yeast. *EMBO J.* **17**, 5525–5528 (1998).
- Critchlow, S. E. & Jackson, S. P. DNA joining: from yeast to man. *Trends Biochem. Sci.* **23**, 394–398 (1998).
- Lewis, L. K., Westmoreland, J. W. & Resnick, M. A. Repair of endonuclease-induced double-strand breaks in *Saccharomyces cerevisiae*: essential role for genes associated with nonhomologous end-joining. *Genetics* **152**, 1513–1529 (1999).
- Boulton, S. J. & Jackson, S. P. Identification of a *Saccharomyces cerevisiae* Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic Acids Res.* **24**, 4639–4648 (1996).
- Astrom, S. U., Okamura, S. M. & Rine, J. Yeast cell-type regulation of DNA repair. *Nature* **397**, 310 (1999).
- Lee, S. E., Paques, F., Sylvan, J. & Haber, J. E. Role of yeast *SIR* genes and mating type in directing DNA double-strand breaks to homologous and non-homologous repair paths. *Curr. Biol.* **9**, 767–770 (1999).
- Worcel, A., Strogatz, S. & Riley, D. Structure of chromatin and the linking number of DNA. *Proc. Natl Acad. Sci. USA* **78**, 1461–1465 (1981).
- Wechsner, M. A., Kladde, M. P., Alfieri, J. A. & Peterson, C. L. Effects of Sin—versions of histone H4 on yeast chromatin structure and function. *EMBO J.* **16**, 2086–2095 (1997).
- Usachenko, S. I., Bavykin, S. G., Gavin, I. M. & Bradbury, E. M. Rearrangement of the histone H2A C-terminal domain in the nucleosome. *Proc. Natl Acad. Sci. USA* **91**, 6845–6849 (1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as a paper copy from the London editorial office of Nature.

Acknowledgements

We thank J. Rouse, F. Winston, N. Lowndes and S. Elledge for providing strains and reagents; and G. Rathbun, N. Lowndes, F. Winston and members of the S.P.J. laboratory, in particular, S. Bell, J. Bradbury and J. Rouse, for helpful discussions.

Correspondence and requests for materials should be addressed to S.P.J. (e-mail: spj13@mole.bio.cam.ac.uk).

Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain

Zhihong Liu*, Chaohong Sun*, Edward T. Olejniczak*, Robert P. Meadows*, Stephen F. Betz*, Thorsten Oost*, Julia Herrmann†, Joe C. Wu† & Stephen W. Fesik*

* Pharmaceutical Discovery Division, Abbott Laboratories, Abbott Park, Illinois 60064, USA

† Biochemistry Department, Idun Pharmaceutical Inc., La Jolla, California 92037, USA

The inhibitor-of-apoptosis proteins (IAPs)¹ regulate programmed cell death by inhibiting members of the caspase family of enzymes^{2–5}. Recently, a mammalian protein called Smac⁶ (also named DIABLO⁷) was identified that binds to the IAPs and promotes caspase activation. Although undefined in the X-ray structure, the amino-terminal residues of Smac are critical for its function^{8,9}. To understand the structural basis for molecular recognition between Smac and the IAPs, we determined the solution structure of the BIR3 domain of X-linked IAP (XIAP) complexed with a functionally active nine-residue peptide derived from the N terminus of Smac. The peptide binds across the third β -strand of the BIR3 domain in an extended conformation with only the first four residues contacting the protein. The complex is stabilized by four intermolecular hydrogen bonds, an electrostatic interaction involving the N terminus of the peptide, and several

hydrophobic interactions. This structural information, along with the binding data from BIR3 and Smac peptide mutants reported here, should aid in the design of small molecules that may be used for the treatment of cancers that overexpress IAPs^{10–12}.

To identify a suitable NMR sample, we compared the binding affinities of full-length Smac, a 20-mer, and a 9-mer Smac peptide to the BIR2 and BIR3 domains of XIAP (Table 1). Although it has been reported that the Smac protein is a dimer and more active than Smac peptides at promoting caspase-3 activation⁹, we find that the Smac protein has the same affinity as the Smac peptides for binding to the BIR3 domain of XIAP. In addition, the Smac peptides and protein bind more tightly to the BIR3 domain (residues 241–356) than to the BIR2 domain (residues 124–240) of XIAP. Thus, we conducted our structural studies on a complex composed of the BIR3 domain of XIAP (Fig. 1a) and a 9-mer peptide (A1'–E9') derived from the N terminus of mature Smac (Fig. 1b) that was able to prevent XIAP from inhibiting caspase-9 (data not shown).

We determined the structure of the XIAP BIR3/Smac peptide complex from a total of 1,487 NMR-derived restraints. The structure for residues 258–346 of the XIAP BIR3 domain and residues 1'–4' of the Smac peptide were well defined by the NMR data (Fig. 2a). In contrast, the carboxy-terminal amino acids of the Smac peptide (residues 5'–9') were disordered.

The structure of the BIR3 domain of XIAP when complexed to the Smac peptide consists of a three-stranded β -sheet, five α -helices, one turn of a 3_{10} helix, and a zinc atom chelated to three cysteines and one histidine (Fig. 2b). It is similar to the structures of the free BIR2¹³ and BIR3 domains¹⁴ of XIAP and survivin^{15–17}. However, two loop regions (residues 276–280 and 308–314) of the XIAP BIR3 domain, which could not be structurally characterized in the free protein¹⁴, become well defined on binding to the Smac peptide.

The Smac peptide adopts an extended conformation with a kink at the proline (residue 3') when bound to the protein, and lies across the third β -strand of the XIAP BIR3 domain (Fig. 2b). Although the

side chain of E314 is not as well-defined as the other residues in the binding site, the carboxyl group of E314 is close to the N terminus of the Smac peptide, suggesting an electrostatic interaction between these moieties. The indole ring of W310 sits underneath the methyl group of A1', consistent with its upfield methyl proton chemical shift (–0.05 p.p.m.). The A1' carbonyl group forms a hydrogen bond with the indole NH of W323, and the NH of V2' hydrogen-bonds to the carbonyl of T308. Additional intermolecular hydrogen bonds occur between the carbonyl of V2' and the NH of T308, and between the NH of I4' and the carbonyl of G306 (Fig. 2c,d). The side chain of V2' interacts with the methyl group of T308, while the proline residue of the peptide contacts the indole ring of W323. I4' of the Smac peptide sits in a channel formed by the hydrophobic portions of K297 and K299 of the BIR3 domain. The remaining residues of the peptide do not show any intermolecular nuclear Overhauser effects (NOEs) with the protein and do not appear to be critical for binding to the BIR3 domain of XIAP.

To identify the amino acids of Smac that are important for complex formation, we synthesized mutant peptides and tested their binding to the BIR3 domain of XIAP using a fluorescence polarization competition assay (Table 1). A peptide in which the N-terminal alanine was replaced with a glycine lost all binding affinity for the protein, suggesting that the hydrophobic interaction between the alanine methyl group and W310 is important. Based on the tight fit of the methyl group into the pocket, larger hydrophobic side chains would not be expected to bind to this site—which is consistent with the inability of a Smac mutant (Ala to Met at the N terminus) to promote caspase-3 activation⁹. Acylation of the N terminus of the Smac peptide also caused a complete loss of binding to the protein (Table 1). To test whether this loss of binding is due to a steric clash from the acyl group or the loss of the positively charged amine at the N terminus, peptides in which the N-terminal amine is substituted by a proton or methyl group (propionic or isobutyric acid versus Ala) were synthesized and tested for binding to the XIAP BIR3 domain. These peptides did not

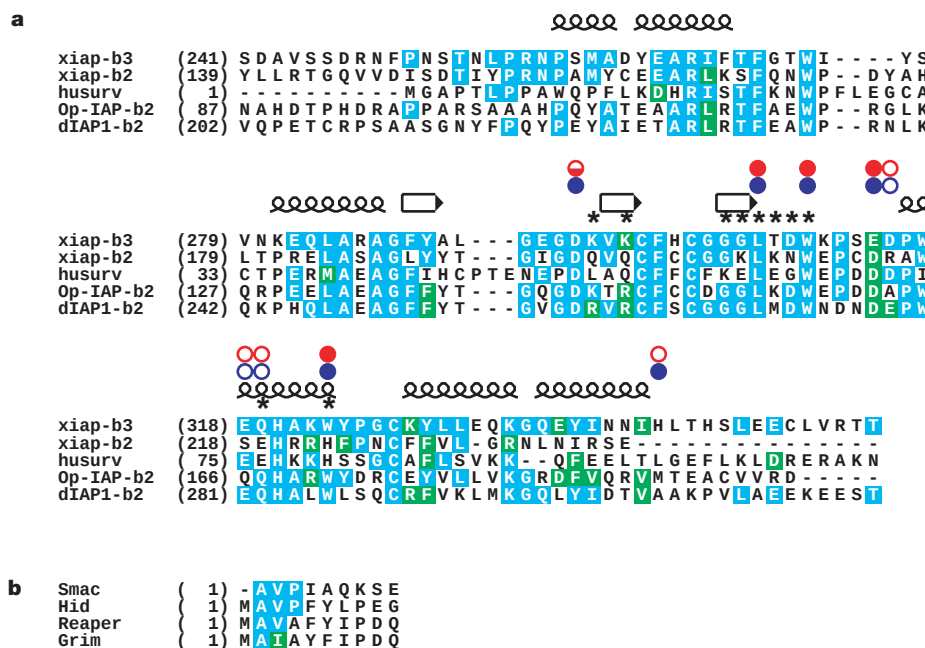


Figure 1 Similarity of IAP BIR domains and N-terminal sequences of IAP binding proteins. **a**, Sequence alignment of the BIR domains and surrounding residues of IAPs that bind to Smac or to Reaper, Hid and Grim. Identical residues are highlighted in blue, and highly homologous residues are highlighted in green. Residues that have intermolecular NOEs to the Smac peptide are indicated with asterisks. The secondary structural elements in the NMR structure of BIR3 are shown above the sequence. Circles denote the residues that

were mutated and tested for caspase-9 inhibition (blue) and Smac peptide binding (red). Filled circles indicate a >10-fold decrease in caspase-9 inhibition or Smac peptide binding, half-filled circles indicate >5-fold decrease, and open circles indicate mutants that have no significant effect. **b**, Sequence alignment of the N-terminal residues of mature Smac, Hid, Reaper and Grim. Residue 1 of mature Smac corresponds to residue 56 of unprocessed Smac.

bind to the protein, suggesting that the positive charge is critical for complex formation. Mutation of the valine, proline or isoleucine of the peptide also caused a loss of binding to the protein, albeit less than A1' substitutions (Table 1). These results can be explained by a loss of hydrophobic interactions when these residues are substituted by an alanine. Amino acids C-terminal to I4' are not important for binding to the protein, because peptides containing mutations of these residues exhibit wild-type affinity. In fact, the 9-mer peptide can be shortened to a 5-mer and still retain full binding affinity to the BIR3 domain of XIAP (Table 1).

We identified the amino acids of the BIR3 domain critical for binding to the Smac peptide by testing mutant proteins for binding to a fluorescein-labelled Smac peptide (Table 2). To ensure that the mutant proteins were folded, $^{15}\text{N}/^1\text{H}$ correlation spectra were recorded, and a well-dispersed spectrum was observed for each protein. As expected, proteins containing alanine mutations of residues distant from the Smac binding site (D315, E318, H343) retain wild-type binding affinity. In contrast, when the hydrophobic amino acids that define the A1' binding pocket (L307, W310) or the

P3' interaction site (W323) were mutated, a significant loss in binding to the Smac peptide was observed. Furthermore, the E314S mutant BIR3 domain lost almost all binding (Table 2), consistent with an electrostatic interaction between E314 and the N-terminal amine of Smac. Despite its close proximity to the Smac peptide, mutation of Q319 caused only a slight loss in binding affinity. Another mutant, D296A, binds ten times more weakly to the Smac peptide and yet is not directly involved in binding to the Smac peptide. This can be explained by the importance of the D296 carboxyl group in stabilizing the structure of the binding site through interactions with the backbone NH of D309 (Fig. 2c).

In general, the ability of the mutant proteins to bind to the Smac peptide matched their ability to inhibit caspase-9 (Table 2, Fig. 1a). These results suggest that Smac and caspase-9 bind to overlapping sites on the BIR3 domain of XIAP, and that Smac acts by freeing caspase-9 from the XIAP/caspase-9 complex. The binding sites for Smac and caspase-9 are not identical, however, as shown by the lack of caspase-9 inhibition observed for H343A which binds very well to the Smac peptide (Table 2). H343 is located at the C-terminal end of

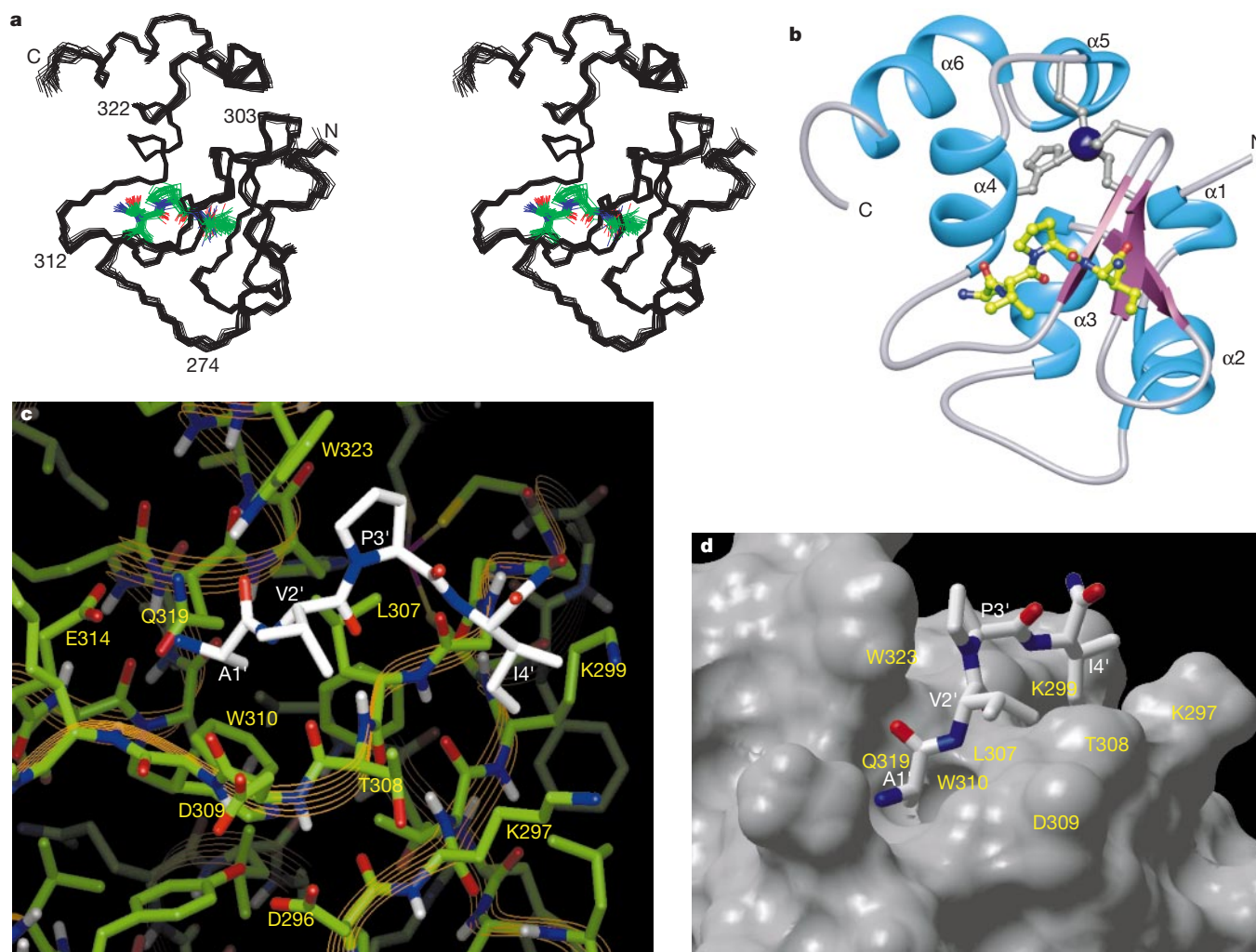


Figure 2 Structure of the XIAP BIR3 domain/Smac peptide complex. **a**, Stereo view of 30 superimposed NMR-derived structures of the BIR3 domain of XIAP (showing residues 258–346) complexed to the Smac peptide (showing residues 1'–4'). The Smac peptide is colour-coded by atom type. The root mean square deviation (r.m.s.d.) to the mean coordinate positions for residues 260–345 of the protein is $0.38 \pm 0.04 \text{ \AA}$ for the backbone atoms (N, C $^{\alpha}$, C') and $0.78 \pm 0.05 \text{ \AA}$ for all heavy atoms. The distance restraint force constant was $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, and no NOE was violated by more than $0.4 \text{ \AA}$. The final energy of the NOE constraints (ENOE) was $24 \pm 13 \text{ kcal mol}^{-1}$. The torsion restraint force constant was $200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$, and no torsion restraint was violated by more

than 5° . Only the covalent geometry terms, NOE, torsion, and repulsive van der Waals terms were employed in the refinement. Even so, a large, negative Lennard–Jones potential energy was observed ($-399 \pm 13 \text{ kcal mol}^{-1}$), indicating good non-bonded geometry. Procheck²⁵ analysis indicated that 97% of the residues are in allowed regions of the Ramachandran map. **b**, Ribbons²⁷ representation of the averaged minimized NMR structure of the XIAP BIR3/Smac peptide complex. **c**, Close-up of the binding site. The carbon atoms of the Smac peptide (A1'–I4') are shown in white. **d**, Surface representation of the binding site of the BIR3 domain bound to the Smac peptide.

the fifth α -helix and is far away from the Smac binding site.

On the basis of primary sequence alignments (Fig. 1a), it is likely that Smac interacts with IAPs in the same manner as shown here for the XIAP BIR3/Smac peptide complex. Not only are the residues in the Smac binding site conserved (for example, L307, W310, E314, W323), but the three-dimensional structures in this region of the BIR3 domain of XIAP—and of the other structurally characterized IAPs that bind to Smac—are also similar. The root mean square deviation (r.m.s.d.) for the comparison¹⁸ of the complexed XIAP BIR3 domain with the free XIAP BIR2 domain¹³ (64 C α atoms) is 1.7 Å and with uncomplexed human survivin¹⁵ (61 C α atoms) is 1.6 Å. Some of these structural differences may be due to conformational changes upon complexation. There are also a few differences in the amino acids that form the Smac binding site in these proteins that probably confer binding specificity. For example, the XIAP BIR2 domain contains bulkier amino acid residues compared to the XIAP BIR3 domain (K206 versus G306, and K208 versus T308) in the region that interacts with P3' and I4' (Fig. 1a). This may explain the differences in binding affinities observed for the P3'A and I4'A peptides to the XIAP BIR2 and BIR3 domains (Table 1). Functional homologues of Smac found in lower organisms (Reaper, Hid, Grim)¹⁹ may also bind to IAPs in a similar fashion, on the basis of their sequence homology at the N terminus (Fig. 1b) which is the critical portion of these proteins for interacting with IAPs¹⁹. Although a Hid peptide with an N-terminal methionine did not bind to the XIAP BIR2 or BIR3 domain, tight binding of a Hid peptide was observed when the N-terminal methionine was removed (Table 1) which probably occurs to the native protein during normal processing in *Drosophila*²⁰. The increase in affinity of

the Hid versus Smac peptide is primarily due to the phenylalanine in the 4' position (Table 1).

The IAPs, such as XIAP and survivin, are widely expressed in human cancers^{10–12} and represent targets for cancer therapy. We consider that the structural information presented here on the functional groups required for the XIAP/Smac interaction should aid in the design of small molecules intended to bind to the IAPs and induce apoptosis in a fashion similar to Smac. These compounds may be useful for treating cancers in which IAPs are overexpressed. □

Methods

Preparation of proteins and peptides

Smac (residues 56–239) was cloned from a human liver complementary DNA library into the vector pET-23b (Novagen) and expressed in the *Escherichia coli* strain BL21(DE3) (Novagen). Recombinant Smac containing a C-terminal histidine tag was purified using a nickel-nitrilotriacetic acid (NTA) column (Qiagen). Electrospray mass spectrometry was used to determine that the N-terminal methionine was quantitatively removed. The BIR2 and BIR3 domains of XIAP were prepared as previously described^{13,14}. Mutants of the XIAP BIR3 domain were prepared using the QuikChange site-directed mutagenesis kit (Stratagene) and coding sequences were confirmed by DNA sequencing. The Smac peptides were obtained from Synpep Corporation (Dublin, California) or synthesized in-house using an Applied Biosystems peptide synthesizer.

NMR spectroscopy

Samples for NMR contained a 1:1 complex of the XIAP BIR3 domain (residues 241–356) and the Smac peptide (residues 1'–9') at a concentration of 1.0 mM each in 10 mM phosphate buffer (pH 7.2), 120 mM NaCl, and 1 mM dithiothreitol. Two samples were used in the study. One sample consisted of uniformly ¹⁵N-¹³C-labelled protein and unlabelled peptide and the other sample contained unlabelled protein and peptide selectively ¹⁵N-¹³C-labelled at A1', V2', I4' and A5'. NMR spectra were acquired at 30 °C on a Bruker 500, 600 or 800 MHz NMR spectrometer. The ¹H, ¹⁵N and ¹³C resonances of the backbone and side chains were assigned using a standard set of double and triple resonance experiments²¹. Distance restraints were derived from ¹⁵N- and ¹³C-resolved 3D NOESY (nuclear Overhauser enhancement spectroscopy) experiments. Slowly exchanging amide protons were identified from a series of ¹H-¹⁵N heteronuclear single quantum coherence spectra after the H₂O buffer was exchanged for a buffer containing ²H₂O.

Structure calculations

In the first step of these calculations, 557 unambiguous NOEs, 36 hydrogen-bond restraints, and 86 ϕ or ψ angular restraints derived from the TALOS program²² were incorporated into a torsion-angle dynamics²³ and simulated annealing protocol²⁴ using the program CNX (MSI Inc., San Diego). A single family of low-energy structures was obtained from these calculations that defined the overall fold. Refinement of these initial structures was accomplished using the ARIA protocol²⁵. New NOE assignments were accepted from ARIA if they agreed with the manually derived NOEs. From 7 iterations, 1,270 unambiguous NOE assignments were derived.

Caspase-9 inhibition assay

Caspase-9 inhibition was measured using a truncated form of the enzyme that lacks the N-terminal CARD domain as previously described¹⁴.

Fluorescence polarization assay

The binding constant of Smac peptides for the BIR2 and BIR3 domains of XIAP was measured using a fluorescence polarization-based competition assay using a Smac-based peptide labelled with 6-carboxyfluorescein succinimidyl ester, FAM, with the sequence AVPIAQKSEK(FAM).

Received 18 September; accepted 7 November 2000.

Table 1 Dissociation constants of Smac peptides for BIR3 and BIR2 domains

Peptide/protein	BIR3 3	BIR2
Mature Smac (56–239)	0.42 ± 0.02	2.3 ± 0.3
AVPIAQKSEPHSLSSSEALMR	0.69 ± 0.05	6.9 ± 0.7
AVPIAQKSE	0.43 ± 0.06	6.0 ± 0.9
GVPIAQKSE	> 1,000	> 1,000
CH ₃ CO-AVPIAQKSE	> 1,000	> 1,000
CH ₃ CH ₂ CO-VPIAQKSE	> 1,000	> 1,000
(CH ₃) ₂ CHCO-VPIAQKSE	> 1,000	> 1,000
AAPIAQKSE	12 ± 2	56 ± 5
AVAIAQKSE	20 ± 4	4.0 ± 0.9
AVPAIAQKSE	34 ± 7	18 ± 2
AVPIAQKSE	1.2 ± 0.4	10 ± 2
AVPIAAKSE	0.43 ± 0.08	3.5 ± 0.2
AVPIAQASE	0.43 ± 0.08	7.1 ± 0.6
AVPIAQKS-NH ₂	0.80 ± 0.03	13 ± 3
AVPIAQK-NH ₂	0.70 ± 0.09	9.4 ± 0.6
AVPIAQ-NH ₂	0.8 ± 0.02	8.9 ± 0.6
AVPIA-NH ₂	0.64 ± 0.07	5.5 ± 0.5
MAVPIFYLPPEGGADDVAS	> 1,000	> 1,000
AVPIFYLPPEG	0.13 ± 0.03	4.9 ± 0.8
AVPIFAQKSE	0.27 ± 0.03	14 ± 3
AVPIYQKSE	0.5 ± 0.10	2.5 ± 0.7

This table shows the effect of N-terminal amino acid changes on the binding affinity of Smac peptides for the BIR3 and BIR2 domains of XIAP. The mutated residues are indicated in bold. K_D, dissociation constants

Table 2 Binding affinity and caspase-9 inhibition of BIR3 mutants

BIR3 mutant	K _D (μM)	IC ₅₀
Wild type	0.74 ± 0.07	16 ± 8 nM
H343A	0.47 ± 0.03	> 1 μM
D315A	0.87 ± 0.05	64 ± 18 nM
E318A	1.6 ± 0.4	31 ± 14 nM
Q319A	2.0 ± 0.8	9 ± 8 nM
D296A	9 ± 2	> 1.7 μM
W323A	42 ± 2	> 2 μM
L307A	63 ± 3	> 2 μM
E314S	210 ± 10	> 2 μM
W310A	> 1,000	> 2 μM

This table shows the effect of mutations on the XIAP BIR3 domain on the binding affinity (K_D) for a fluorescein-labelled Smac peptide and *in vitro* inhibition of caspase-9 (IC₅₀).

10. Tamm, I. *et al.* Expression and prognostic significance of IAP-family genes in human cancers and myeloid leukemias. *Clin. Cancer Res.* **6**, 1796–1803 (2000).
11. Ambrosini, G., Adida, C. & Altieri, D. C. A novel anti-apoptosis gene, survivin, expressed in cancer and lymphoma. *Nature Med.* **3**, 917–217 (1997).
12. LaCasse, E. C., Baird, S., Korneluk, R. G. & MacKenzie, A. E. The inhibitors of apoptosis (IAPs) and their emerging role in cancer. *Oncogene* **17**, 3247–3259 (1998).
13. Sun, C. *et al.* NMR structure and mutagenesis of the inhibitor-of-apoptosis protein XIAP. *Nature* **401**, 818–822 (1999).
14. Sun, C. *et al.* NMR structure and mutagenesis of the third BIR domain of the inhibitor of apoptosis protein XIAP. *J. Biol. Chem.* **275**, 33777–33781 (1998).
15. Verdecia, M. A. *et al.* Structure of the human anti-apoptotic protein survivin reveals a dimeric arrangement. *Nature Struct. Biol.* **7**, 602–608 (2000).
16. Muchmore, S. W. *et al.* Crystal structure and mutagenic analysis of the inhibitor-of-apoptosis protein survivin. *Mol. Cell* **6**, 173–182 (2000).
17. Chantlat, L. *et al.* Crystal structure of human survivin reveals a bow-tie-shaped dimer with two unusual α -helical extensions. *Mol. Cell* **6**, 183–189 (2000).
18. Kleywegt, G. J. & Jones, T. A. in *From First Map to Final Model* (eds Bailey, S., Hubbard, R. & Waller, D.) 59–66 (Daresbury Laboratory, Daresbury, UK, 1994).
19. Vucic, D., Kaiser, W. J. & Miller, L. K. Inhibitor of apoptosis proteins physically interact with and block apoptosis induced by *Drosophila* proteins Hid and Grim. *Mol. Cell Biol.* **18**, 3300–3309 (1998).
20. Cutforth, T. & Gaul, U. A methionine aminopeptidase and putative regulator of translation initiation is required for cell growth and patterning in *Drosophila*. *Mech. Dev.* **82**, 23–28 (1999).
21. Clore, G. M. & Gronenborn, A. M. Multidimensional heteronuclear magnetic resonance of proteins. *Methods Enzymol.* **239**, 349–363 (1994).
22. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
23. Stein, E. G., Rice, L. M. & Brünger, A. T. Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J. Magn. Reson. B* **124**, 154–164 (1997).
24. Nilges, M., Gronenborn, A. M., Brünger, A. T. & Clore, G. M. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng.* **2**, 27–38 (1988).
25. Nilges, M., Macias, M. J., O'Donoghue, S. I. & Oschkinat, H. Automated NOESY interpretation with ambiguous distance restraints: The refined NMR solution structure of the pleckstrin homology domain from β -spectrin. *J. Mol. Biol.* **269**, 408–422 (1997).
26. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
27. Carson, M. J. Ribbon models of macromolecules. *J. Mol. Graphics* **5**, 103–106 (1987).

Acknowledgements

We thank L. Barrett and P. Richardson for help in the synthesis of some of the peptides, L. Miesbauer for the mass spectrometry data, and J. Noel and co-workers for the coordinates of human survivin.

Correspondence and requests for materials should be addressed to S.W.F. (e-mail: Stephen.Fesik@abbott.com). Coordinates for the NMR structure of the XIAP BIR3/Smac peptide complex have been deposited in the Brookhaven Protein Data Bank under accession code 1G3F.

Structural basis of IAP recognition by Smac/DIABLO

Geng Wu[†], Jijie Chai[†], Tomeka L. Suber[‡], Jia-Wei Wu^{*}, Chunying Du[§], Xiaodong Wang[§] & Yigong Shi^{*}

^{*} Department of Molecular Biology, Princeton University, Lewis Thomas Laboratory, Washington Road, Princeton, New Jersey 08544, USA
[‡] 310 Aycock Dorm, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514, USA

[§] Department of Biochemistry and Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA

[†] These authors contributed equally to this work

Apoptosis is an essential process in the development and homeostasis of all metazoans^{1–4}. The inhibitor-of-apoptosis (IAP) proteins suppress cell death by inhibiting the activity of caspases; this inhibition is performed by the zinc-binding BIR domains^{5,6} of the IAP proteins. The mitochondrial protein Smac/DIABLO promotes apoptosis by eliminating the inhibitory effect of IAPs through physical interactions^{7–9}. Amino-terminal sequences in Smac/

DIABLO are required for this function, as mutation of the very first amino acid leads to loss of interaction with IAPs and concomitant loss of Smac/DIABLO function⁹. Here we report the high-resolution crystal structure of Smac/DIABLO complexed with the third BIR domain (BIR3) of XIAP. Our results show that the N-terminal four residues (Ala-Val-Pro-Ile) in Smac/DIABLO recognize a surface groove on BIR3, with the first residue Ala binding a hydrophobic pocket and making five hydrogen bonds to neighbouring residues on BIR3. These observations provide a structural explanation for the roles of the Smac N terminus as well as the conserved N-terminal sequences in the *Drosophila* proteins Hid/Grim/Reaper. In conjunction with other observations, our results reveal how Smac may relieve IAP inhibition of caspase-9 activity. In addition to explaining a number of biological observations, our structural analysis identifies potential targets for drug screening.

All members of the IAP family contain at least one BIR (baculoviral IAP repeat) motif and many contain three. Recent experiments indicate that different BIR domains may exhibit distinct functions. The second BIR domain (BIR2) of XIAP is a potent inhibitor for caspase-3, whereas the third BIR domain of XIAP (XIAP-BIR3) primarily targets the active caspase-9^{10–13}. Smac is synthesized as a precursor molecule of 239 amino acids; the N-terminal 55 residues serve as the mitochondria targeting sequence that is removed after import⁷. The wild-type Smac protein (residues 1–184) forms a dimer in solution and interacts with both the BIR2 and BIR3 domains of XIAP⁹. Point mutations at the dimeric interface, such as Phe33→Asp, completely inactivate dimer formation⁹. Such monomeric mutants can no longer interact with XIAP-BIR2, but they retain binding to the BIR3 domain⁹. A missense mutation of the N-terminal Ala residue to Met (Ala1→Met) in wild-type Smac abolishes binding to both the BIR2 and BIR3 domains of XIAP and results in complete loss of Smac function⁹. In support of its critical role, a short 7-residue peptide derived from the Smac N-terminus was able to promote the activation of procaspase-3⁹, raising the possibility of using small peptides or peptide mimics to treat cancer cells.

To provide a structural basis for IAP recognition by Smac, we crystallized the wild-type Smac in complex with either the BIR2 or the BIR3 domain of XIAP. But these crystals did not diffract X-rays well. To facilitate crystallization, we reconstituted binary complexes using a monomeric Smac protein (with the missense mutation Phe33→Asp). Crystals of this mutant Smac with XIAP-BIR3 (residues 238–358) were obtained in two different conditions, neither of which disrupted interactions between Smac and XIAP-BIR3 in solution. We have determined these two crystal structures of the binary complex at 2.0 and 2.6 Å (see Methods and Supplementary Information). There are two identical complexes in each asymmetric unit in each of the two crystal forms. Although the packing interactions are different, two sets of conserved interaction interfaces between Smac and XIAP-BIR3 are present in these two crystal forms, with 0.95 Å root-mean-square-deviation (r.m.s.d.) for all aligned C α atoms (Fig. 1a). Because of these conserved features, we limit our discussions to the 2.0 Å structure (Fig. 1b).

In the complex, the Smac protomer is an elongated three-helix bundle (Fig. 1), very similar to the structure of Smac by itself (0.7 Å r.m.s.d. for aligned C α atoms)⁹. XIAP-BIR3 consists of six α -helices, a three-stranded β -sheet, and a zinc atom chelated by three Cys and one His residues (Cys 300, Cys 303, His 320 and Cys 327) (Fig. 1). The XIAP-BIR3 structure closely resemble those of other BIR domains¹⁴, with approximately 1.22 and 1.38 Å r.m.s.d. to the aligned C α atoms of survivin^{15,16} and cIAP1-BIR2¹², respectively. In the crystals, Smac recognizes XIAP-BIR3 with two interfaces. First, the N-terminal four residues in Smac, Ala1-Val2-Pro3-Ile4, bind a surface groove on BIR3 formed by the strand β 3, the helix α 3 and the intervening loop. Two of these four residues, Val2 and Pro3, form a short anti-parallel β -strand with the three-stranded β -sheet

of BIR3 (Figs 1a and 2a). Second, the helices H2 and H3 in Smac contact the BIR3 residues surrounding the helix $\alpha 1$ (Fig. 1a). The N-terminal residues and the H2–H3 helices in Smac that contact the same BIR3 domain are from two different Smac protomers (Fig. 1b). The N-terminus of one Smac protomer extends out and recognizes a BIR3 domain that interacts with another Smac protomer through the second interface (Fig. 1b). This arrangement is in agreement with our structure-based prediction⁹.

The N-terminal four residues of Smac pack into a surface groove on XIAP-BIR3, resulting in the burial of 892 Å² exposed surface area (Fig. 2b). The recognition specificity is achieved through a combination of hydrogen-bond interactions and van der Waals contacts (Fig. 2a). A total of eight inter- and three intra-molecular hydrogen bonds support the binding of the Smac tetrapeptide (Ala1–Val2–Pro3–Ile4) in the surface groove on BIR3. Three intermolecular contacts between the backbone groups of Val2/Ile4 in Smac and Gly306/Thr308 in BIR3 allow the formation of a four-stranded antiparallel β -sheet (Fig. 2c). The remaining hydrogen bonds constitute an intricate network surrounding the N-terminal residue Ala1. At the centre of the network, the amino group of Ala 1 donates three hydrogen bonds to Glu 314 and Gln 319 whereas its carbonyl group makes two additional contacts to Gln 319 and Trp 323

(Fig. 2c). At the periphery of the network, three intra-molecular contacts further buttress the interactions (Fig. 2c).

In addition to the hydrogen-bond network, van der Waals contacts also appear to be important in stabilizing the interactions between the Smac tetrapeptide and the BIR3 surface groove. The methyl group of Ala 1 fits tightly in a hydrophobic pocket formed by the side chains of Leu 307, Trp 310 and Gln 319 (Fig. 2c). Stereochemical parameters indicate that replacement of Ala 1 by any other residue except Gly will cause steric hindrance in this pocket, probably weakening binding and abolishing hydrogen bonds by the amino and carbonyl groups of Ala 1. This observation explains the finding that the mutation Ala1→Met in Smac completely eliminated interaction with the BIR domains⁹. The absolute requirement for Ala as the N-terminal residue is also consistent with the observation that the mutant Smac-del4 retained weak interaction with the BIR domains⁹, as removal of the first four residues in Smac leaves Ala 5 as the N-terminal residue. We predict that Ala 1 cannot be replaced by Gly because of the associated entropic penalty and loss of van der Waals interactions, although we have not tested this prediction.

The bulky, flat aromatic side chain of Trp 323 is important in forming the surface groove on BIR3 (Fig. 2c). Both Val 2 and Pro 3

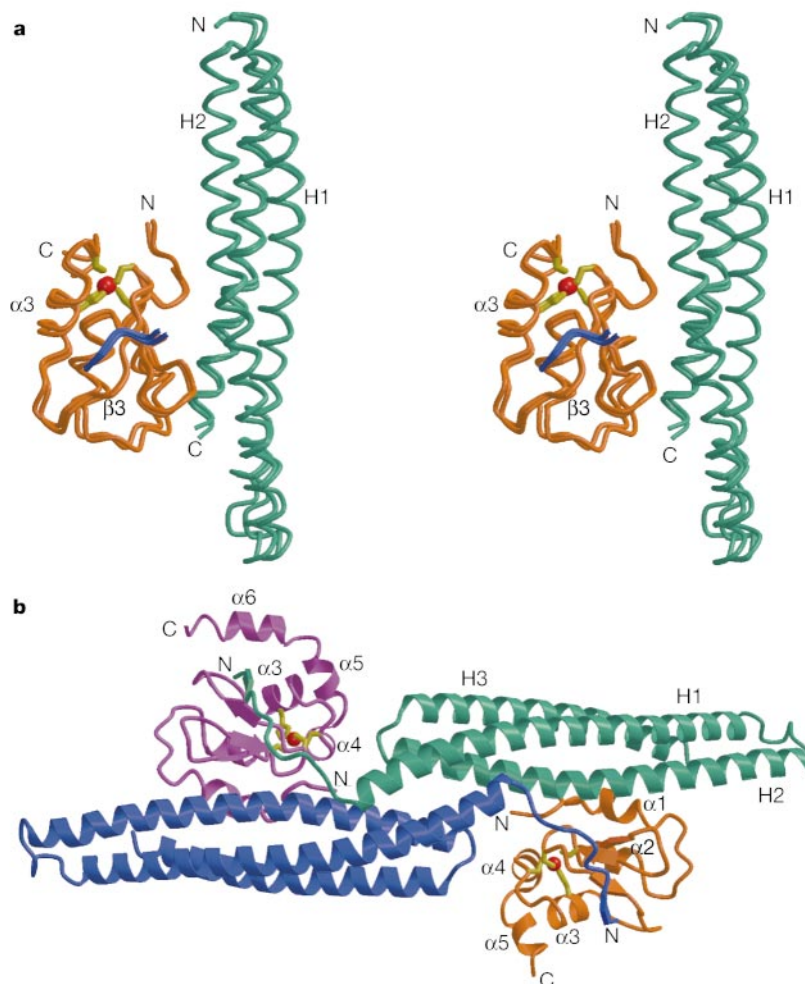


Figure 1 Schematic representation of the Smac(Phe33→Asp)/XIAP-BIR3 structures. **a**, Stereo view of superimposition of the two structures from two different crystal forms. Smac and XIAP-BIR3 are coloured green and orange, respectively. The bound Smac N-terminal peptide is shown in blue. The zinc atom is highlighted in red whereas its chelating residues are shown in yellow. **b**, Schematic representation of the complete structure in one asymmetric unit. Two Smac protomers are coloured green and blue,

respectively. Two BIR3 domains are shown in orange and pink, respectively. Some secondary structural elements are labelled. In the crystals, each Smac protomer uses its extended N terminus to bind a BIR3 domain whereas it interacts with a second BIR3 domain with a different interface. All figures were prepared with MOLSCRIPT²⁶ and GRASP²⁷.

maintain multiple van der Waals interactions with Trp 323, while Pro 3 makes additional contact to Tyr 324 (Fig. 2c). In addition, the side chain of Ile 4 interacts with Leu 292, Gly 306, and the aliphatic side chains of Lys 297 and Lys 299. Of these residues, Gly 306 appears to be particularly important, because the absence of a side chain at this position makes these interactions possible (Fig. 2c). Although Ala 5 and Gln 6 in Smac are ordered in the crystals, they do not make important interactions to the BIR3 domain; nor are they involved in binding the surface groove.

Among the N-terminal four residues, Ala 1 makes the largest contribution to the specific recognition of XIAP. The next three residues, Val 2, Pro 3 and Ile 4, are all hydrophobic, and they interact with the hydrophobic residues lining the BIR3 surface groove. These residues also contribute to the appropriate positioning of Ala 1 in the pocket (Fig. 2c); however, their identity does not appear to be as critical as that of the N-terminal Ala 1.

Our structural analysis has identified several residues in XIAP-BIR3 that mediate crucial interactions with Smac. To confirm these structural observations, we created nine missense mutations on BIR3 and examined their interaction with Smac using purified recombinant proteins. Consistent with their important roles in binding the Smac N-terminus, point mutation to Ala or Arg for any of the three residues Glu 314, Gln 319 and Trp 323 significantly reduced or abolished interaction with the wild-type Smac protein, presumably because of the disruption of binding to the Smac tetrapeptide (data not shown). In contrast, three control mutations outside the peptide-binding surface, Glu282→Ala, Glu282→Arg and Arg286→Glu, had no detectable effect on the interaction between XIAP-BIR3 and Smac (data not shown).

In *Drosophila*, three proteins, Hid, Grim and Reaper, appear to be the functional homologues of the mammalian Smac⁹. These three proteins appear to act upstream of the *Drosophila* IAP, DIAP1, and

physically interact with DIAP1 to relieve its inhibitory effect on caspase activation^{17,18}. Although previous sequence alignment between Smac and the *Drosophila* proteins failed to reveal significant homology, our results showing a Smac tetrapeptide binding to XIAP motivated us to re-examine the N-terminal sequences of Hid/Grim/Reaper. This analysis revealed notable similarity in the N termini of these proteins (Fig. 2d). All three *Drosophila* proteins begin with Ala; both Reaper and Hid have Val in the second position whereas Grim replaces Val with a conserved Ile. This sequence conservation strongly suggests that the *Drosophila* proteins Hid/Grim/Reaper may interact with DIAP1 in a similar fashion, and that the N-terminal sequences from Hid/Grim/Reaper may recognize the same surface groove. This analysis also predicts that (1) the initiating Met residue in Hid/Grim/Reaper is removed by a methionyl peptidase in *Drosophila* as is the case in *Escherichia coli*^{9,19}, and (2) other pro-apoptotic proteins bearing N-terminal homology to the Ala-Val-Pro-Ile sequence may remain to be discovered.

In addition to the peptide-binding groove, we also observed a second interaction interface between Smac and the BIR3 domain of XIAP (Fig. 2e). Residues around the helix $\alpha 1$ in BIR3 pack against the middle portion of helices H2 and H3 in Smac. This interface consists of seven hydrogen bonds and two patches of van der Waals interactions, with over 2,000 Å² of buried surface area (Fig. 2e). The hydrogen bonds cluster in a small area, where Glu 99 and Thr 100 in Smac make six contacts to Asn 259, Ser 261 and Arg 258 (Fig. 2e). In addition, Arg 85 on helix H2 donates one hydrogen bond to the side chain of Thr 274 (Fig. 2e). Although van der Waals interactions are scattered throughout the large interface (Fig. 2e), there are two prominent patches. First, Phe 270 on helix $\alpha 1$ of XIAP-BIR3 extends out into a hydrophobic pocket formed by the side chains of Met 88, Leu 153, Ala 154, Gln 157 and Glu 150 (Fig. 2e). Second, Leu 96 on helix H2 makes multiple contacts to the side chains of Met 262 and

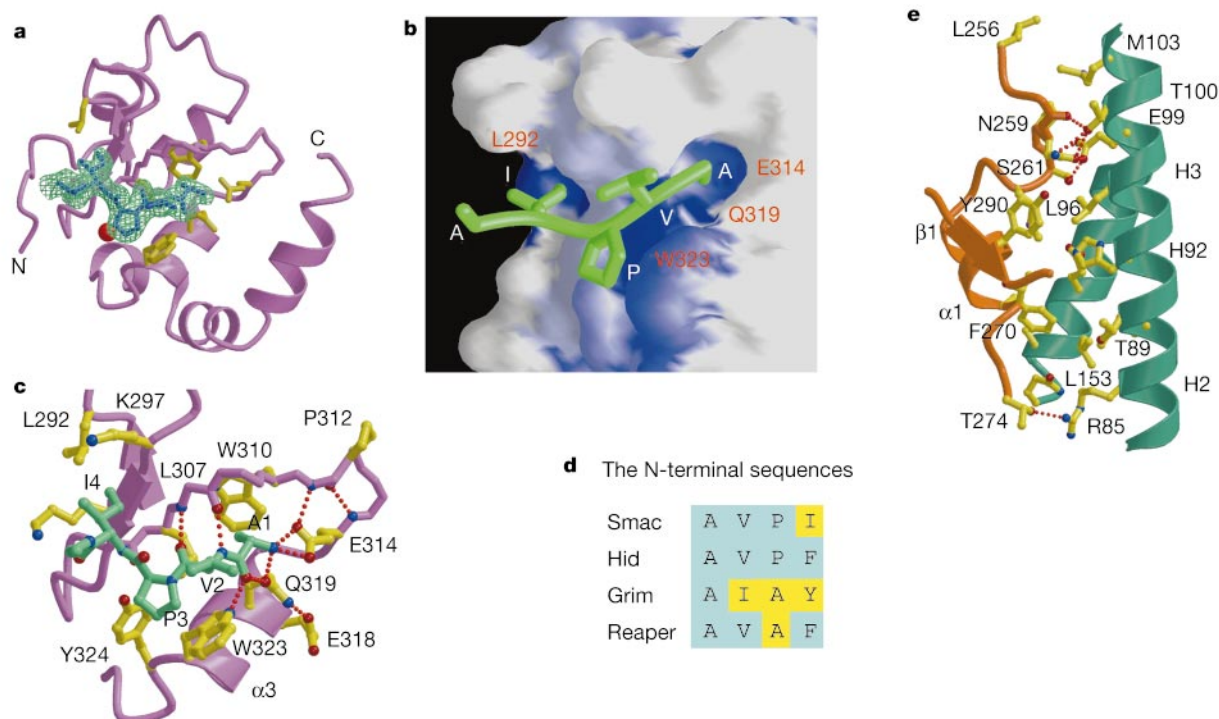


Figure 2 Binding interface between Smac and XIAP-BIR3. **a**, Overall view of the interaction between Smac N-terminal residues and XIAP-BIR3. Smac is coloured blue and the interface residues from BIR3 are highlighted in yellow. The zinc atom is shown in red. The $F_o - F_c$ electron density (omit map), shown in green, was contoured at 2.5σ and calculated with simulated annealing using XPLOR²⁴ with the omission of the Smac N-terminal seven residues. **b**, Close-up view of the binding groove on BIR3 for the Smac N-terminus. The blue and white colours represent the most and least hydrophobic surfaces, respectively. **c**, Close-up view of the interaction between the Smac N terminus and the surface groove on BIR3. Smac and BIR3 are coloured green and pink, respectively. Hydrogen bonds are represented by red dashed lines. Oxygen and nitrogen atoms are shown as red and blue balls, respectively. **d**, Alignment of the N-terminal four amino acids of Smac with those from the *Drosophila* proteins Hid, Grim and Reaper. **e**, Close-up view of a second interface between Smac and XIAP-BIR3. Smac and BIR3 are coloured green and orange, respectively.

respectively. The N-terminal five residues of Smac are shown in green. **c**, Close-up view of the interaction between the Smac N terminus and the surface groove on BIR3. Smac and BIR3 are coloured green and pink, respectively. Hydrogen bonds are represented by red dashed lines. Oxygen and nitrogen atoms are shown as red and blue balls, respectively. **d**, Alignment of the N-terminal four amino acids of Smac with those from the *Drosophila* proteins Hid, Grim and Reaper. **e**, Close-up view of a second interface between Smac and XIAP-BIR3. Smac and BIR3 are coloured green and orange, respectively.

Tyr 290 (Fig. 2e). Although the buried surface area is relatively large for this interaction interface, shape complementarity between Smac and BIR3 is not optimal. Compared to the binding of the Smac N-terminal tetrapeptide to the surface groove on BIR3, this interface is likely to play a minor role. This analysis is consistent with the observation that N-terminal deletion mutants of Smac failed to bind XIAP-BIR3⁹.

Wild-type Smac interacts with the BIR2 and BIR3 of XIAP but does not bind the BIR1 domain⁹. To explain this observation, we compared the primary sequences of the BIR domains from XIAP and cIAP-1. Among the many residues that line the BIR3 surface groove, five appear to be mediating critical contacts with Smac. Leu 307 and Trp 310 are involved in pocket formation for the side chain of Ala 1 of Smac; Glu 314 hydrogen-bonds to Ala 1; Trp 323 interacts with Val 2 and Pro 3 and hydrogen-bonds to Ala 1; Gly 306 allows docking of Ile 4. Among these five residues, Trp 310 is invariant; Leu 307 and Glu 314 are highly conserved among all BIR domains. The replacement of Gly 306 by other residues in BIR2 and BIR1 probably disrupts tight packing of Ile 4 of Smac against residues in the BIR domain. More importantly, Trp 323 is replaced by His in the BIR2 domain and Val/Leu in the BIR1 domain, which may alter the lining of the surface groove. The replacement of Trp 323 by His in BIR2 is likely to be less deleterious than its replacement by Val/Leu; this is because His may retain the hydrogen bond contact to the carbonyl group of Ala 1 in Smac (Fig. 2c). These analyses are consistent with the observations that

wild-type Smac does not interact with BIR1 and that monomeric Smac mutants fail to bind BIR2⁹.

The BIR3 domain of XIAP is primarily responsible for the inhibition of caspase-9 activity^{10,13}. How does Smac relieve this inhibition? Both published observations and current analysis suggest a model of mutual exclusion of caspase-9 and Smac for binding to the BIR3 domain. Three mutations in XIAP-BIR3, Glu314→Ser, Trp310→Ala and His343→Ala, abolish inhibition of caspase-9 activity¹³, suggesting that the affected residues may directly contact caspase-9 and inhibit its activity. Indeed, the three affected residues are close to each other and two of them are directly involved in binding Smac tetrapeptide (Fig. 3a). In the structure, Trp 310 forms an important part of the pocket that accommodates Ala 1 in Smac, suggesting that binding by the Smac N terminus may evict a binding moiety of caspase-9 from the pocket. In addition, the amino group of Ala1 in Smac makes a pair of charge-stabilized hydrogen bonds to the side chain of Glu 314 in XIAP, probably competing with the interaction between Glu 314 and caspase-9. His343→Ala affects a residue that abuts the peptide-binding pocket, further supporting the mutual exclusion model (Fig. 3a).

Our structural analysis also provides a plausible explanation for gain-of-function (GOF) mutations in DIAP1. All five reported GOF mutations^{18,20} appear to disrupt the peptide-binding groove and impair binding by the *Drosophila* proteins Hid/Grim/Reaper (Fig. 3a). For example, Gly269→Ser affects a residue that corresponds to G306 in XIAP-BIR3; Val85→Met and Gly88→Asp (or Gly88→Ser) affect residues that correspond to Gly 305 and Thr 308 in XIAP, respectively. These residues are either in the groove (Gly 306 and Thr 308) or in the vicinity of the groove (Gly 305). The fifth GOF mutation, Pro105→Ser, affects a corresponding residue (Pro 325) in XIAP that makes a sharp turn and allows Trp 323 to line the groove. We suggest that these five mutations impair binding of DIAP-1 to Hid/Grim/Reaper without significantly affecting binding to *Drosophila* caspases (Fig. 3a).

Although a Smac monomeric mutant (Phe33→Asp) was used for crystallization, there are two complexes in each asymmetric unit in the crystals (Fig. 1b). The two Smac protomers pack against each other using part of the wild-type dimeric interface (Fig. 1b). Docking the XIAP-BIR3 domain onto the wild-type Smac dimer structure using the BIR3:Smac interface observed here places the BIR3 domain underneath the arch-shaped Smac molecule (Fig. 3b). This arrangement also places two BIR3 domains next to each other (Fig. 3b). This model for the complex between wild-type Smac and BIR3 is consistent with available biochemical data. For example, XIAP-BIR2 does not interact with the monomeric mutants of Smac, suggesting that BIR2:BIR2 interactions may be required for stable binding of BIR2 to the wild-type dimeric Smac⁹.

Our high-resolution crystal structure of a complex between Smac and the BIR3 domain of XIAP reveals a peptide-binding groove on the surface of BIR3 that is important in IAP function. This peptide-binding groove is lined with both hydrophobic and negatively charged residues, and appears to be a promising drug target. Future experiments need to be directed at identification and optimization of a high-affinity drug that is capable of promoting apoptosis in cancer cells. □

Methods

Site-directed mutagenesis and protein preparation

Point mutations were generated by PCR, and the identities of individual clones were verified by sequencing. Recombinant XIAP-BIR3 (residues 238–358) were overexpressed as GST-fusion proteins using pGEX-2T (Pharmacia). The mutant Smac protein (Phe33Asp, residues 1–162) was overexpressed in *E. coli* strain BL21(DE3) using a pET-3d vector (Novagen). Selenomethionyl Smac (Phe33Asp) was expressed in *E. coli* B834(DE3) (Novagen) in M9 minimal medium supplemented with 50 mg l⁻¹ selenomethionine. Protein purification was performed as described⁹. The concentration of the final complex was approximately 15 mg ml⁻¹. For interaction assays, both wild-type and mutant Smac were overexpressed in *E. coli* strain BL21(DE3) as C-terminally 9-Histidine-tagged proteins using a pET-15b vector (Novagen).

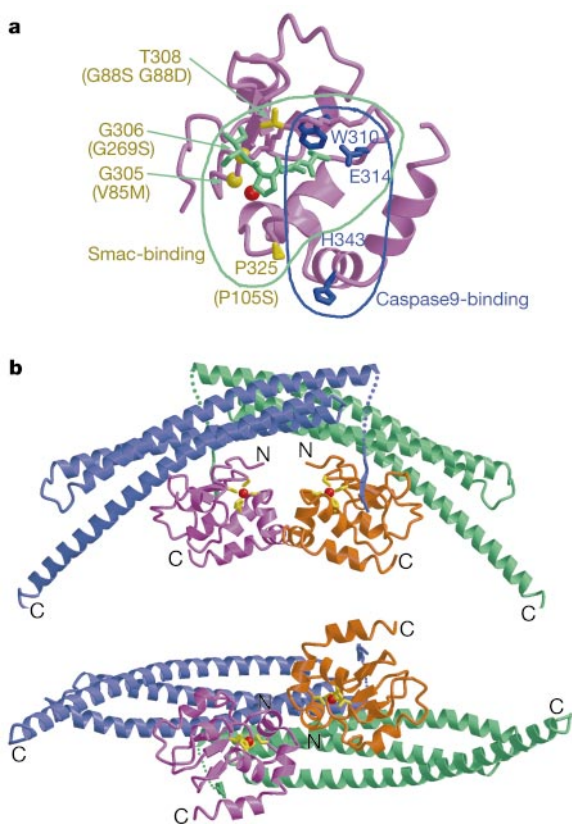


Figure 3 Proposed mechanism for the relief of XIAP inhibition of caspase-9 by Smac. **a**, The surface on XIAP-BIR3 that binds to caspase-9 overlaps with the Smac-binding groove. XIAP-BIR3 is primarily responsible for inhibiting caspase-9^{10,13}. Mutation of the three residues, W310, E314 and H343 (coloured blue), prevents caspase-9 inhibition by XIAP-BIR3¹³. Four residues, T308, G305, G306 and P325, are shown in yellow. Mutation of the corresponding residues in the *Drosophila* protein DIAP1 (shown in parentheses) leads to a gain-of-function phenotype^{18,20}. **b**, Proposed structure of a wild-type Smac/XIAP-BIR3 complex (see text). Two perpendicular views are shown.

In vitro interaction assay

Interaction between Smac and XIAP-BIR3 was examined by GST-mediated pull-down assays. Approximately 0.4 mg of a wild-type or mutant BIR3 fragment was bound to 200 μ l of glutathione resin as a GST-fusion protein and incubated with 0.5 mg of wild-type or mutant Smac at room temperature. After extensive washing with an assay buffer containing 25 mM Tris, pH 8.0, 150 mM NaCl, and 2 mM dithiothreitol (DTT), the complex was eluted with 5 mM reduced glutathione and visualized by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) with Coomassie staining.

Crystallization and data collection

Crystals were grown by the hanging-drop vapour-diffusion method by mixing the Smac/XIAP-BIR3 complex (15 mg ml⁻¹) with an equal volume of reservoir solution. At 4 °C, the reservoir contained 100 mM MES buffer, pH 6.5, 12% isopropanol (v/v), 200 mM sodium citrate, and 10 mM DTT. At 23 °C, the reservoir contained 100 mM citrate buffer, pH 5.5, 5% PEG 4000 and 10% isopropanol. Crystals appeared after 1–4 days and reached a maximum size over a period of 1–3 weeks. Both crystals are in the triclinic space group P1 and contain two complexes in each unit cell. However, the unit cell dimensions are significantly different. Crystals grown at 4 °C have $a = 47.0$ Å, $b = 54.6$ Å, $c = 74.2$ Å, $\alpha = 93.0^\circ$, $\beta = 101.1^\circ$, and $\gamma = 94.9^\circ$; crystals grown at 23 °C have $a = 48.1$ Å, $b = 52.9$ Å, $c = 67.1$ Å, $\alpha = 100.0^\circ$, $\beta = 104.1^\circ$, and $\gamma = 94.0^\circ$. Diffraction data were primarily collected using an R-AXIS-IV imaging plate detector mounted on a Rigaku 200HB generator. Derivatives were obtained by soaking crystals in reservoir buffer containing 20% glycerol (v/v) and heavy atoms. The concentration and soaking time for mercury thimerosal were 1 mM and 20 hours, respectively. To collect data at -170 °C, crystals were equilibrated in a cryoprotectant buffer containing reservoir buffer plus 20% glycerol (v/v) and were flash frozen in a cold nitrogen stream. The high-resolution native data were collected at the NSLS beamline X4A.

Structure determination

We focused on the crystals grown at 4 °C first, because of the availability of a high-quality selenomethionine derivative. The first six selenium positions were determined using SOLVE²¹ and further refined using MLPHARE²². The positions of the other four selenium atoms and mercury thimerosal were identified using difference Fourier methods. Initial MIR phases calculated with the program MLPHARE²² had a mean figure of merit of 0.404 to 3.0 Å resolution, and were improved with solvent flattening and histogram matching using the program DM²². The electron density for the Smac N-terminal four residues was very clear in the experimental map. A model was built into MIR electron density with the program O²³ and refined at 2.6 Å resolution by simulated annealing using the program XPLOR²⁴. Non-crystallography symmetry (NCS) constraints were applied to all refinement cycles except the last one. The two refined complexes (R factor 24% and free R 28.5%) are nearly identical to one another, and contain Smac residues 1–6 and 13–157, XIAP residues 251–343, and 58 ordered water molecules. Residues 7–12 and 158–162 in Smac and the terminal residues in XIAP (238–250 and 344–358) have no electron density in the maps, and we presume that these regions are disordered in the crystals. After completion of the refinement for the crystals grown at 4 °C, the atomic coordinates of BIR3 and Smac were used individually to obtain molecular replacement solutions for the crystals grown at 23 °C, using AMoRe²⁵. The solutions were combined in O²³. Rigid body refinement by XPLOR²⁴ confirmed the correctness of the solutions. This model was further refined at 2.0 Å resolution by simulated annealing using the program XPLOR²⁴, followed by model building in O²³. NCS was helpful for the initial cycles of refinement. The final refined model, with an R factor of 22.7% and free R of 26.8%, contains two complexes. One complex contains Smac 1–157 and XIAP 256–357, and the other complex contains Smac 1–157 and XIAP 256–343. Except for the C-terminal 14 residues of XIAP in the first complex, which is disordered in the second one, the structures of these two complexes are identical to one another.

Received 18 September; accepted 9 November 2000.

1. Steller, H. Mechanisms and genes of cellular suicide. *Science* **267**, 1445–1449 (1995).
2. Jacobson, M. D., Weil, M. & Raff, M. C. Programmed cell death in animal development. *Cell* **88**, 347–354 (1997).
3. Hengartner, M. O. Programmed cell death in invertebrates. *Curr. Opin. Genet. Dev.* **6**, 34–38 (1996).
4. Horvitz, H. R. Genetic control of programmed cell death in the nematode *Caenorhabditis elegans*. *Cancer Res.* **59**, 1701–1706 (1999).
5. Deveraux, Q. L. & Reed, J. C. IAP family proteins—suppressors of apoptosis. *Genes Dev.* **13**, 239–252 (1999).
6. Miller, L. K. An exegesis of IAPs: salvation and surprises from BIR motifs. *Trends Cell Biol.* **9**, 323–328 (1999).
7. Du, C., Fang, M., Li, Y. & Wang, X. Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation during apoptosis. *Cell* **102**, 33–42 (2000).
8. Verhagen, A. M. *et al.* Identification of DIABLO, a mammalian protein that promotes apoptosis by binding to and antagonizing IAP proteins. *Cell* **102**, 43–53 (2000).
9. Chai, J. *et al.* Structural and biochemical basis of apoptotic activation by Smac/DIABLO. *Nature* **406**, 855–862 (2000).
10. Deveraux, Q. L. *et al.* Cleavage of human inhibitor of apoptosis protein XIAP results in fragments with distinct specificities for caspases. *EMBO J.* **18**, 5242–5251 (1999).
11. Takahashi, R. *et al.* A single BIR domain of XIAP sufficient for inhibiting caspases. *J. Biol. Chem.* **273**, 7787–7790 (1998).
12. Sun, C. *et al.* NMR structure and mutagenesis of the inhibitor-of-apoptosis protein XIAP. *Nature* **401**, 818–822 (1999).
13. Sun, C. *et al.* NMR structure and mutagenesis of the third BIR domain of the inhibitor of apoptosis

protein XIAP. *J. Biol. Chem.* **275**, 33777–33781 (2000).

14. Shi, Y. Survivin structure: crystal unclear. *Nature Struct. Biol.* **7**, 620–623 (2000).
15. Verdecia, M. A. *et al.* Structure of the human anti-apoptotic protein survivin reveals a dimeric arrangement. *Nature Struct. Biol.* **7**, 602–608 (2000).
16. Chantalat, L. *et al.* Crystal structure of human survivin reveals a bow-tie-shaped dimer with two unusual alpha helical extensions. *Mol. Cell* **6**, 183–189 (2000).
17. Wang, S., Hawkins, C., Yoo, S., Muller, H.-A. & Hay, B. The *Drosophila* caspase inhibitor DIAP1 is essential for cell survival and is negatively regulated by HID. *Cell* **98**, 453–463 (1999).
18. Goyal, L., McCall, K., Agapite, J., Hartwig, E. & Steller, H. Induction of apoptosis by *Drosophila* reaper, hid and grim through inhibition of IAP function. *EMBO J.* **19**, 589–597 (2000).
19. Hirel, P.-H., Schmitter, J.-M., Dessen, P., Fayat, G. & Blanquet, S. Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl Acad. Sci. USA* **86**, 8247–8251 (1989).
20. Lisi, S., Mazzon, I. & White, K. Diverse domains of THREAD/DIAP1 are required to inhibit apoptosis induced by REAPER and HID in *Drosophila*. *Genetics* **154**, 669–678 (2000).
21. Terwilliger, T. C. & Berendzen, J. Correlated phasing of multiple isomorphous replacement data. *Acta Crystallogr. D* **52**, 749–757 (1996).
22. Collaborative Computational Project. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
23. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
24. Brunger, A. T. *X-PLOR, a System for Crystallography and NMR* (Yale Univ. Press, New Haven, Connecticut, 1991).
25. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
26. Klauis, P. J. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
27. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296 (1991).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank C. Ogata for help at the NSLS-X4A beamline; F. Hughson for critically reading the manuscript; S. Kyin for peptide sequencing, DNA synthesis and mass spectroscopy; and N. Hunt for secretarial assistance. This research is supported by start-up funds from Princeton University (Y.S.) and the Howard Hughes Medical Institute (X.W.). Y.S. is a Searle Scholar and a Rita Allen Scholar.

Correspondence and requests for materials should be addressed to Y.S. (e-mail: yshi@molbio.princeton.edu). Atomic coordinates for the Smac/BIR3 complex have been deposited at the Protein Data Bank with accession number 1G73.

correction

Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit

Richard H. R. Hahnloser, Rahul Sarpeshkar, Misha A. Mahowald, Rodney J. Douglas & H. Sebastian Seung

Nature **405**, 947–951 (2000).

There were two errors in the reference list of this paper. Reference 22 should have been:

Sharpshkar, R. Analog versus digital: extrapolating from electronics to neurobiology. *Neural Comput.* **10**, 1601–1638 (1998).

Also, the following reference was omitted but should have been cited along with refs 5–9 at the top of page 948.

Douglas R. J., Koch, C., Mahowald, M. A., Martin, K. A. C. & Suarez, H. Recurrent excitation in neocortical circuits. *Science* **269**, 981–985 (1995).